1976

An analysis of white students' evaluations of black instructors in predominantly white colleges and universities

Earl Glenn Yarbrough

Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Part of the Higher Education Administration Commons, and the Higher Education and Teaching Commons

Recommended Citation
Yarbrough, Earl Glenn, 'An analysis of white students' evaluations of black instructors in predominantly white colleges and universities " (1976). Retrospective Theses and Dissertations. 5721.
https://lib.dr.iastate.edu/rtd/5721

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or “target” for pages apparently lacking from the document photographed is “Missing Page(s)”. If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.

2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in “sectioning” the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.

4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from “photographs” if essential to the understanding of the dissertation. Silver prints of “photographs” may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

Xerox University Microfilms
300 North Zeib Road
Ann Arbor, Michigan 48106
YARBROUGH, Earl Glenn, 1946-
AN ANALYSIS OF WHITE STUDENTS' EVALUATIONS
OF BLACK INSTRUCTORS IN PREDOMINANTLY
WHITE COLLEGES AND UNIVERSITIES.

Iowa State University, Ph.D., 1976
Education, higher

Xerox University Microfilms, Ann Arbor, Michigan 48106
An analysis of white students' evaluations of black instructors in predominantly white colleges and universities

by

Earl Glenn Yarbrough

A Dissertation Submitted to the Graduate Faculty in Partial Fulfillment of The Requirements for the Degree of DOCTOR OF PHILOSOPHY

Major: Industrial Education

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1976
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>ABSTRACT</th>
<th>vi</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHAPTER I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Problem of the Study</td>
<td>6</td>
</tr>
<tr>
<td>Purpose of the Study</td>
<td>6</td>
</tr>
<tr>
<td>Need of the Study</td>
<td>7</td>
</tr>
<tr>
<td>Hypotheses of the Study</td>
<td>12</td>
</tr>
<tr>
<td>Research Hypothesis I</td>
<td>12</td>
</tr>
<tr>
<td>Statistical Hypothesis I</td>
<td>12</td>
</tr>
<tr>
<td>Research Hypothesis II</td>
<td>12</td>
</tr>
<tr>
<td>Statistical Hypothesis II</td>
<td>12</td>
</tr>
<tr>
<td>Research Hypothesis III</td>
<td>13</td>
</tr>
<tr>
<td>Statistical Hypothesis III</td>
<td>13</td>
</tr>
<tr>
<td>Research Hypothesis IV</td>
<td>13</td>
</tr>
<tr>
<td>Statistical Hypothesis IV</td>
<td>13</td>
</tr>
<tr>
<td>Research Hypothesis V</td>
<td>13</td>
</tr>
<tr>
<td>Statistical Hypothesis V</td>
<td>13</td>
</tr>
<tr>
<td>Assumptions of the Study</td>
<td>14</td>
</tr>
<tr>
<td>Limitations of the Study</td>
<td>14</td>
</tr>
<tr>
<td>Procedure of the Study</td>
<td>15</td>
</tr>
<tr>
<td>CHAPTER II. REVIEW OF THE LITERATURE</td>
<td>19</td>
</tr>
<tr>
<td>Student Evaluations in Higher Education</td>
<td>19</td>
</tr>
<tr>
<td>Sex, Class Level and Black-White Differences</td>
<td>23</td>
</tr>
<tr>
<td>Sex differences</td>
<td>23</td>
</tr>
<tr>
<td>Class level differences</td>
<td>25</td>
</tr>
<tr>
<td>Black-white differences</td>
<td>28</td>
</tr>
<tr>
<td>Theory, Reliability, Validity and Use of Student Evaluations</td>
<td>33</td>
</tr>
<tr>
<td>Theoretical base</td>
<td>33</td>
</tr>
<tr>
<td>Reliability and validity</td>
<td>36</td>
</tr>
<tr>
<td>Use of evaluations</td>
<td>40</td>
</tr>
<tr>
<td>Blacks in Higher Education</td>
<td>44</td>
</tr>
</tbody>
</table>
### Chapter Summary

**CHAPTER III. RESEARCH DESIGN**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Instructor Rating Form</td>
<td>49</td>
</tr>
<tr>
<td>Review existing forms</td>
<td>50</td>
</tr>
<tr>
<td>Rating scale</td>
<td>57</td>
</tr>
<tr>
<td>Development of the rating form</td>
<td>58</td>
</tr>
<tr>
<td>Video Tape Production</td>
<td>60</td>
</tr>
<tr>
<td>Selection of lecturers (instructors)</td>
<td>61</td>
</tr>
<tr>
<td>Selection of script</td>
<td>62</td>
</tr>
<tr>
<td>Methods used in taping</td>
<td>64</td>
</tr>
<tr>
<td>Selection of experts</td>
<td>64</td>
</tr>
<tr>
<td>Methods used to review tapes</td>
<td>65</td>
</tr>
<tr>
<td>Analysis of expert scores</td>
<td>66</td>
</tr>
<tr>
<td>Treatment</td>
<td>68</td>
</tr>
<tr>
<td>Description of the population</td>
<td>68</td>
</tr>
<tr>
<td>Selection of the sample</td>
<td>69</td>
</tr>
<tr>
<td>Pilot study</td>
<td>71</td>
</tr>
<tr>
<td>Instructor contacts</td>
<td>72</td>
</tr>
<tr>
<td>Administer treatment</td>
<td>73</td>
</tr>
<tr>
<td>Analysis of data</td>
<td>75</td>
</tr>
</tbody>
</table>

**Chapter Summary**

**CHAPTER IV. FINDINGS**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classes Utilized</td>
<td>77</td>
</tr>
<tr>
<td>Test of Assumptions</td>
<td>79</td>
</tr>
<tr>
<td>Analysis of Data</td>
<td>80</td>
</tr>
<tr>
<td>Research Hypothesis I</td>
<td>81</td>
</tr>
<tr>
<td>Research Hypothesis II</td>
<td>82</td>
</tr>
<tr>
<td>Research Hypothesis III</td>
<td>83</td>
</tr>
<tr>
<td>Research Hypothesis IV</td>
<td>83</td>
</tr>
<tr>
<td>Research Hypothesis V</td>
<td>85</td>
</tr>
</tbody>
</table>
# Chapter Summary

## CHAPTER V. SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary</td>
<td>87</td>
</tr>
<tr>
<td>Conclusions</td>
<td>92</td>
</tr>
<tr>
<td>Research Hypothesis I</td>
<td>92</td>
</tr>
<tr>
<td>Research Hypothesis II</td>
<td>94</td>
</tr>
<tr>
<td>Research Hypothesis III</td>
<td>94</td>
</tr>
<tr>
<td>Research Hypothesis IV</td>
<td>95</td>
</tr>
<tr>
<td>Research Hypothesis V</td>
<td>96</td>
</tr>
<tr>
<td>Restatement of the problem</td>
<td>96</td>
</tr>
<tr>
<td>Recommendations</td>
<td>97</td>
</tr>
</tbody>
</table>

### BIBLIOGRAPHY

- Page 100

### ACKNOWLEDGMENTS

- Page 110

### APPENDIX A: LETTER TO DR. W. J. McKEACHIE

- Page 112

### APPENDIX B: LETTER FROM DR. W. J. McKEACHIE

- Page 114

### APPENDIX C: IOWA STATE UNIVERSITY INSTRUCTOR RATING FORM

- Page 116

### APPENDIX D: INSTRUCTOR RATING FORM

- Page 118

### APPENDIX E: LECTURE SCRIPTS - PARTS I AND II

- Page 120

### APPENDIX F: LETTER FROM DR. D. J. ZAFFARANO

- Page 136
LIST OF TABLES

Table                                      Page
1. Instructor order of presentation on each tape 16
2. $2 \times 3 \times 2$ factorial design with a repeat measure on instructor 17
3. Mean scores of each expert for each tape presentation 67
4. Analysis of variance summary table for experts ratings 67
5. Total number of males and females in each class rating instructors per level and science area 78
6. Mean scores for each class used in the study 79
7. The analysis of variance summary table for black and white instructor ratings 80
8. Analysis of variance table relating to Hypothesis I 81
9. Analysis of variance table relating to Hypothesis II 82
10. Analysis of variance table relating to Hypothesis III 83
11. Analysis of variance table relating to Hypothesis IV 84
12. Analysis of variance table relating to Hypothesis V 85
ABSTRACT

The study was designed to determine if skin color makes a difference in instructor ratings by predominantly white classes in a predominantly white university. The basis for such an investigation is contingent upon two factors occurring in higher education today: (1) the increased emphasis upon student evaluation of faculty and (2) the increased efforts to employ more black faculty at predominantly white colleges and universities. As a result, the black instructor hired by the administration of a predominantly white university must be evaluated by white students. Therefore, it is essential to determine if all faculty are rated equally, regardless of one's skin color.

In addition to determining if instructor skin color made a difference in student evaluations, students' sex and class level (under division, upper division and graduate division) were considered as factors that might effect black and white faculty ratings. Therefore, a 2 x 3 x 2 factorial design with a repeat measure was employed.

The study was conducted using two separate video tapes. The two tapes were divided into two parts. The white instructor presented part 1 of tape I and part 2 of tape II while the black instructor presented part 2 of tape I and part 1 of tape II.
The subject matter of the two tapes was completely the same. Half of the sample (classes) were shown tape I and the other half tape II. This was necessary to control for the effect of either instructor appearing first. An attempt was made to control all other relevant variables except skin color. The tapes were evaluated by a team of experts as being equal in presentation, using the instructor rating form developed for the study.

The tapes were shown to a total of twelve randomly selected classes at Iowa State University: four under division, four upper division, and four graduate division classes. Male and female evaluations within each class were treated separately. The students were told to rate the instructors based upon the instructors' performance. Care was taken not to alert the students to the real reason the evaluation was being conducted. The data was analyzed, using an analysis of variance procedure. The analysis indicated a significant difference on the instructor dimension of the study: \( F (1, 18) = 54.28, p < .05 \). An inspection of the overall mean scores for each instructor indicated that the black instructor was rated higher than the white instructor. For the sex and class level dimension there was reported no significant differences with \( F (1, 18) = .31, p > .05 \) and \( F (2, 18) = .75, p > .05 \) respectively. In addition, there were no significant
interactions between any of the three main effects.

The results of the instructor dimension of the study was contrary to the commonly held view that discrimination if it occurred would be directed toward the black. The author suggests several reasons why such a difference may have occurred and list several recommendations for future research related to the topic.
CHAPTER I. INTRODUCTION

Prejudice and discrimination have become major concerns of the American people. Although the two terms are closely related, they have different meanings. The Living Webster Encyclopedic Dictionary (1975) defines prejudice as:

An opinion, judgment, or evaluation, favorable or more often unfavorable, conceived without proof or competent evidence, but based on what seems valid to one's own mind; a bias against a race, creed, group or the like; the holding of such feelings (p. 751).

The same dictionary defines discrimination as:

The act of discriminating or differentiating, or the resulting state; differentiating; the making of a difference in particular cases, as in favor of or against a person, particularly when influenced by race or creed rather than individual merit (p. 268).

As prejudice may be categorized as "a bias against a race, creed, group or the like", the result of that bias may cause some individuals to discriminate "in favor of or against a person". Therefore, to understand discrimination, a theoretical base for prejudice is helpful. Throughout this study only racial prejudice and discrimination will be discussed.

Several authors have attempted to explain prejudice based upon social-psychological theory. Oliver Cromwell Cox, viewed prejudice as a result of the economic structure. This "exploitation" theory as it is sometimes
referred to, deals with man's manipulation of other men.

Cox (1948) stated:

Race prejudice . . . is a social attitude propagated among the public by an exploiting class for the purpose of stigmatizing some group as inferior so that the exploitation of either the group itself or its resources may both be justified (p. 393).

The theory Cox proposed has been adopted by several blacks who feel they are the victims of the American economic system.

Another theory of significant quality, was propagated by Sigmund Freud. Freud saw prejudice as a form of projection. Paul Roazen (1968) in reviewing Sigmund Freud's work explained projection:

To project, means to treat an internal state as though it were in the external world. For example, people have for centuries negatively projected onto minorities, like Jews, their own disguised fears and impulses (inferiority, lasciviousness, greediness, etc.). Negroes in America have served a similar psychological function for objectifying a slightly different set of anxieties (p. 68).

The concept of projection has also been used by a number of leading sociologists. Two of these sociologists, Putney and Putney (1964) explained their projection theory:

It may become a general custom to alienate certain self-potential and to project it onto a particular group of people - the scapegoat of a society. The sum of such projections constitutes the stereotype of this despised group. Such stereotypes
are defended ardently against all evidence which contradicts them (p. 42).

Another explanation of prejudice is based upon the concept of personality. T. W. Adorno and associates (1950) explained prejudice as a personality trait. The source of prejudice is traced to childhood and the child's relationship with his parents. The parent's influence upon the child makes it impossible for him to tolerate ambiguity. The child is taught that everything is either right or wrong, with little or no exceptions.

Another theory of importance was the frustration-aggression theory of prejudice. Harry L. Miller and Roger R. Woock (1973) explained the theory:

The frustration-aggression theory traces frustration in modern society to a number of causes. It may be the result of a constitutional illness or physical condition. It may come from family relations, from other sources nearby in the community, or from political events rather far removed from the individual. One common human response to all of these frustrations is to direct them on to some other target. This psychological device is called displacement (p. 210).

The theory of displacement can best be illustrated using the military chain-of-command. The Sergeant displaces his anger upon the corporal, the corporal does the same with the private, the private screams at his best friend, and so the cycle continues. This theory has also been associated with the term scapegoat - the scapegoat being the object of an individual's displaced frustration.
Society was the basis for an additional theory of prejudice. The theory was built around the individual and his conformity to societal norms. The theory is explained by Miller and Woock (1973):

... this theory might account for patterns of prejudice toward blacks in the American South. These negative feelings and attitudes are an important part of the social values of the 'Southern way of life' (p. 209).

It is likely that one or all the theories discussed are important in understanding race prejudice. S. P. Adinaraya (1964) lists and explains thirteen causes for color (race) prejudice:

1. Fear
2. Jealousy
3. A sense of unfamiliarity
4. A sense of superiority
5. Ignorance
6. Frustration
7. Conditioning
8. Ethnocentrism
9. Antistic thinking
10. Philosophic despair
11. Pathological sources
12. Visibility
13. Sexual riots (pp. 6-26)
It is likely, that one or all the theories discussed might be used to explain racial prejudice. A person's economic status, background, geographic location may all be factors that affect an individual's concepts of others. The Freudian theory that explains the Nazi's contempt of the Jewish people during World War II may not be adequate to explain the American white's view of his black neighbor.

With such theoretical knowledge of prejudice, one can better understand discrimination. Unlike prejudice, discrimination can be regulated and controlled by the courts to improve the relationship between the races. Prejudice is a hard-to-touch concept of the human mind; discrimination is a form of behavior, and behavior is, of the two, obviously more subject to regulation.

Even though racial discrimination is a form of behavior, it is nonetheless difficult to detect. Often, what one may judge to be racial discrimination may be action prompted by motives other than prejudices.

Blacks, whites, Japanese and Mexicans often judge other humans by the color of their skin and nothing else. Discrimination can be observed in many phases of our everyday lives. On the streets, in and out of the ghettos, between employee-
employer and even in the churches. Although there are numerous places one can observe racial discrimination, this proposal will focus upon discrimination and its effect upon the college environment. With the increasing emphasis upon minority-majority relationships in higher education, the researcher has evaluated one aspect of that relationship - the aspect of discrimination and its effect upon instructor evaluation.

Problem of the Study

The problem of the study was to ascertain if skin color made a difference in instructor ratings by classes in a predominantly white university.

Purpose of the Study

The purpose of the study was three-fold:

1. To provide the black educator in the predominantly white university with information that he can use to review his student's evaluation of his performance.

2. To provide administrators in higher education with data that will be helpful in reviewing student's evaluations of black instructors.
3. To provide a means of detecting racial discrimination and to initiate programs to eliminate discrimination.

Need of the Study

The need for this study was established in view of two trends occurring in higher education. (1) The predominantly white institutions' efforts to attract and employ more black instructors and (2) the increased emphasis on instructor accountability through student evaluations. These trends have brought about the need for research to determine if black instructors have been fairly evaluated by predominantly white classes.

With increased pressure from the federal government, minority relations have become a major topic of many institutions of higher education. During the late sixties, attempts at racial harmony were most apparent. Harold T. Johnson (1968) wrote:

Little doubt remains among the citizenry that racial integration has been permanently declared the policy of the three branches of the federal government. An increasing number of citizens have realized that the future of the nation depends upon unity in its economic, social, political, and spiritual life. Lack of racial harmony with its concomitant problems presents the greatest threat to national unity at this time (p. 147).

To further emphasize the point, Walter J. Ducey in his
The article, "Equal Employment Opportunity Comes To The Campus" reviews the federal laws that have led to the present employment practices used by the universities. Ducey (1974) stated:

The law has indeed changed hiring and promotional practices in business and industry and can be expected to do the same in educational institutions opening up to minorities and women a far greater range of employment opportunities than have been open in the past, no matter how good the intentions of the institutions may have been (p. 1).

The increased recruitment of black instructors in white institutions was the topic of David M. Rafky's (1972) research on black scholars and job opportunities:

I pursued the matter of job invitation by asking "How many unsolicited job offers have you had within the past year?" Whites report 1.5 mean offers compared to 3.1 for the blacks for the academic year 1968-1969. This is, therefore, a period of reverse discrimination in which blacks already in the academic profession are sought out by predominantly white colleges and universities (p. 256).

These authors are just a few of the many that have written about blacks in higher education. Being a qualified minority in America today means greater employment opportunities in the colleges and universities.

As more blacks were being added to the faculties of predominantly white universities, blacks faced new and challenging situations. One of these situations concerned the method used by the university administration
to evaluate the instructional staff. As students were demanding and receiving a larger voice in determining their educational needs, the faculty and staff were being asked to become more accountable. One such method of evaluation was student ratings of faculty. Therefore, students' evaluation of college instructors was the second area important to understanding the need for the research conducted.

Wilbert J. McKeachie (1969a), in an article reviewing student ratings of faculty members, wrote:

In recent months, a growing number of students and faculty members have evinced interest in making more effective use of student evaluations of courses and teachers in higher education (p. 438).

In the same article, McKeachie (1969a) notes that student's ratings of university faculty is rather new:

The students should serve as the "experts" in evaluating the effectiveness of their instruction is a relatively new and revolutionary idea in the field of higher education. No one has doubted that students have opinions about the quality of instruction they receive, but only within the past four decades have these ideas been systematically gathered (p. 438).

McKeachie has outlined the rapid growth and development of student evaluations in higher education.

Other authors have also noted the increased interest in faculty evaluation by students. Two authors in particular, H. Richard Smock and Terence J. Crooks (1973)
noted that several separate groups were advocating evaluation of college teaching for several different reasons:

There has been rapidly developing pressure in recent years to provide evaluative data on college teaching: pressure from teaching faculty, from students, and from administrators. The major reasons for the pressure seem to be that teaching faculty want information which will aid them in improving their instruction, students want information to guide them in course and instructor selection, and administrators want information to guide them in pay and promotion decisions (p. 577).

Not every faculty member, student and administrator would agree with Smock and Crooks, but what was important was that pressure was being placed upon those involved in higher education to evaluate the college instructor.

Thomas M. Sherman and John L. Winstead confirmed that student evaluations were used in a number of different ways. The two authors noted, as Smock and Crooks did, that salary and promotions were ways student evaluations of instruction had been used. Sherman and Winstead (1975) wrote:

The more common proposed uses for student ratings include providing course-end feedback for instructors, evaluating teaching competence for promotion purposes and salary adjustments and providing the student body with information for selecting courses (or more properly, instructors) (p. 34).

Although there were a number of reasons for adopting a student rating system, Sherman and Winstead identify
several that did effect the college professors future. As a great deal depended upon these student ratings, it was imperative that they be a "true" reflection of the instructors professional abilities in and out of the classroom.

From the discussion presented thus far, it has been documented that two emerging factors have gained support in higher education. The two are: (1) the increased number of black professors hired by "white" universities and (2) the rapid growth and use of student evaluations to rate faculty. As a result, the black instructor hired by a predominantly white university had to be evaluated by white students. Consequently, the black instructor was placed in an uncertain position of determining if his evaluation by his students were a reflection of his efforts and abilities or were they based upon other less tangible variables. In light of the discussion on prejudice, and its relationship to discrimination, the black instructors rating may have been founded upon racial discrimination. Therefore, it was essential that research be conducted to determine if racial discrimination was a factor that effected black instructors ratings by predominantly white students.
Hypotheses of the Study

Research Hypothesis I

It was hypothesized that there is no significant difference in overall student rating of black and white instructors on the instructor rating form.

Statistical Hypothesis I

\[ H_0: \mu_B = \mu_W \quad \mu_B = \text{mean for black instructor} \]
\[ H_a: \mu_B \neq \mu_W \quad \mu_W = \text{mean for white instructor} \]
\[ \alpha = .05 \]

Research Hypothesis II

It was hypothesized that there is no significant difference between graduate division, upper division and lower division class ratings of black and white instructors on the instructor rating form.

Statistical Hypothesis II

\[ H_0: \mu_{un} = \mu_{up} = \mu_{gr} \quad \mu_{un} = \text{mean of under division} \]
\[ H_a: \text{At least one pair of means is not equal} \]
\[ \mu_{up} = \text{mean of upper division} \]
\[ \mu_{gr} = \text{mean of graduate division} \]
\[ \alpha = .05 \]
Research Hypothesis III

It was hypothesized that there is no significant difference between male and female ratings of black and white instructor on the instructor rating form.

Statistical Hypothesis III

\[ \begin{align*}
\text{Ho: } \mu_F &= \mu_M & \mu_F &= \text{mean of female group} \\
\text{Ha: } \mu_F &\neq \mu_M & \mu_M &= \text{mean of male group} \\
\alpha &= .05
\end{align*} \]

Research Hypothesis IV

It was hypothesized that there is no first-order interaction between instructor x class, instructor x sex and class x sex on the instructor rating form.

Statistical Hypothesis IV

\[ \begin{align*}
\text{Ho: There was no first-order interactions} \\
\text{Ha: There was (were) first-order interaction(s)}
\end{align*} \]

Research Hypothesis V

It was hypothesized that there is no second-order interaction between instructor x class x sex on the instructor rating form.

Statistical Hypothesis V

\[ \begin{align*}
\text{Ho: There was no second-order interaction} \\
\text{Ha: There was a second-order interaction}
\end{align*} \]
Assumptions of the Study

For the purpose of this study, the following assumptions were made:

1. Ratings of the two video tape lessons by the experts were evaluated as being equal.
2. Iowa State University represented a typical predominantly white university.
3. The sample was representative of white students in predominantly white colleges and universities.
4. The Hawthorne effect, if it existed, was equally distributed.
5. The instructor evaluation instrument used provided valid information to test each hypothesis.

Limitations of the Study

The study was conducted under the following limitations:

1. The student sample for this study was limited to a random selection of classes in which students at Iowa State University were enrolled.
2. The student's evaluation were based upon a brief exposure to white and black instructors.
3. The study was limited to one university student population.
Procedure of the Study

The following stepwise procedure was followed in the completion of the study:

1. Selected instructor evaluation instrument forms were reviewed.
2. Existing rating scales used on instructor evaluation instruments were reviewed.
3. An instructor rating form for use in this study was developed.
4. One black and one white instructor was chosen based upon instructor selection criteria.
5. The topic selected for writing the scripts for the tape production was DRUGS IN THE SCHOOLS. The topic was thought to be of interest to all students regardless of age or sex.
6. Methods used in taping were considered.
7. Three experts were chosen to evaluate the tape production. All three experts had had some experience with student evaluations in higher education.
8. Video tapes were made of the two instructors. Each instructor presented parts one and two but not on the same tape, i.e., the white instructor presented part one on tape one and part two on
tape two. The black instructor presented part two of tape one and part one of tape two (see Table 1).

<table>
<thead>
<tr>
<th></th>
<th>Tape I</th>
<th>Tape II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part 1</td>
<td>White instructor</td>
<td>Black instructor</td>
</tr>
<tr>
<td>Part 2</td>
<td>Black instructor</td>
<td>White instructor</td>
</tr>
</tbody>
</table>

9. All attempts were made to equate all relevant variables except for skin color.
10. The completed tapes were evaluated by the experts.
11. The experts scores were analyzed using a 3 x 4 randomized block design with one replicate per cell (Dayton 1970, p. 163).
12. Iowa State University was identified as the population used in the study.
13. Four lower division classes, four upper division classes and four graduate division classes were selected from a stratified random sample of all university classes. The randomly selected classes had to have (1) a 20-80 ratio of men to women or women to men, and (2) two of all four classes selected from each division were of a
social science nature while the other two were of a physical science nature. If any student was enrolled in two or more of the classes selected, only that student's first ratings were used.

14. A pilot study was conducted in order that the presentation of material was clear and precise.

15. The instructors of the classes selected were contacted.

16. The treatment was administered to the selected classes.

17. A 2 x 3 x 2 factorial was used in the design of the study (see Table 2).

Table 2. 2 x 3 x 2 factorial design with a repeat measure on instructor\(^a\)

<table>
<thead>
<tr>
<th>Class Level Variable (A)</th>
<th>Lower Division</th>
<th>Upper Division</th>
<th>Graduate Division</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex Division Variables (B): Male Female</td>
<td>Male Female</td>
<td>Male Female</td>
<td>Male Female</td>
</tr>
<tr>
<td>black: G(_1) G(_2) G(_3) G(_4) G(_4) G(_6)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructor Treatment variable (C):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>white: G(_1) G(_2) G(_3) G(_4) G(_5) G(_6)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)G = a group of four randomly selected classes.

18. Each class was shown part one and part two of either tape one or two. The order-of-treatment (c) was planned so that two of the
four classes representing the group were shown
the black instructor first while the other
two classes were shown the white instructor
first.

19. Upon the completion of viewing each instructor,
the subjects evaluated that instructor using
the evaluation form provided.

20. Data was collected and statistically tested using
the analysis of variance procedure.

21. A summary, with conclusions and recommendations
completed the research project.
CHAPTER II. REVIEW OF THE LITERATURE

This chapter will focus upon two areas directly related to this study: (1) student evaluation of faculty and (2) black faculty in higher education. The review of literature is divided into four parts: Student evaluations in higher education; sex, class level and black-white differences; theory, validity, reliability, and use of student evaluations; and blacks in higher education.

Student Evaluations in Higher Education

The idea that students should be used to evaluate faculty is a relatively new concept in higher education (McKeachie, 1969b). The concept, although relatively new, has produced several studies that have attempted to answer many of the questions posed by concerned faculty. Although there seems to be some negative criticism of student evaluations, (Gage 1961, Bryant 1967 and Kerlinger 1971), most educators support models of teacher evaluation based upon student ratings. A number of authors (Morton 1961, Werdell 1967, and Renner 1967), have indicated support for student evaluations. Although all vary in the extent and use of student evaluations, there is agreement that such evaluations are valuable in improving college teaching. The use of student evaluations have
also been on the increase. The results of a study involving 500 liberal arts colleges indicated that formal student evaluations of faculty had increased from 11% in 1966 to 29% in 1974 (Seldin and Wakin, 1975).

In addition to support for student evaluations of faculty, many authors have attempted to answer some of the questions concerning the evaluation. Some of the more typical questions were: Does the personality of the instructor make a difference in student rating of his teaching ability? Does an easy grader rate higher than the more stringent grader? Are instructors of larger classes rated lower than instructors of small classes? The list of questions is enormous. As a review of student evaluations in higher education, a number of studies will be noted.

Shapiro and Stein (1972) in reviewing two different types of admission policies at an urban university found as a secondary aspect of their study a striking relationship between the frequency of testing and teacher evaluations. The researchers indicated that as the number of exams given in a class increased, the student evaluation of the instructor improved. The results of this study seem to indicate that the personality of the instructor is of less importance than what he or she does in his or her classroom. In this case, the frequency of test
given. The results are somewhat contradictory to what Knapp (1962) indicated in his review of literature on student ratings. Knapp concluded that the instructor himself, who he is, rather than what he does in the classroom is more important in student ratings of faculty.

In another study, Aleamoni, Yimer and Mahan (1972) wanted to know if folklore about a teacher made a difference in his student ratings. The results were based upon student ratings made by two groups of graduate students enrolled in an educational statistics class at the University of Illinois. The first group made ratings during the 1967-68 school year and the second group during the 1968-69 school year. The same instructor taught both classes. The student ratings of the two groups "indicated that students do not build a folklore about a course based upon the course presented one year earlier" (p. 613). Therefore, an instructor's rating based upon his present class are not influenced by former students opinions.

One of the trends developing in higher education during the sixties was an increase in the use of faculty research activity as an indicator of a faculty member's ability (Knapp, 1962). Grant (1971) conducted a study to determine if faculty allocation of time effected his
or her students course evaluations. As a result, the researcher indicated that "there was some evidence that as faculty time allocated to research and writing increased, student ratings of courses decreased" (p. 1). In another study (McDaniel and Feldhusen, 1971) that compared the relationship of authorship to instructional effectiveness, the "instructors whom the students regard as most effective are those who write no books - or who limit their roles as paper and article writers to secondary authorship . . ." (p. 27). The McDaniel and Feldhusen findings were based upon a sample of 4,484 college students.

In another study, Fittante and Powell (1974) investigated the relationship of classroom verbal behavior and student ratings. From the 26 classrooms included in the study, the authors found a strong relationship between classroom interaction and the students ratings of the instructor in terms of motivation and stimulation.

The studies presented thus far are representative of many that have attempted to answer some of the questions concerning student evaluations of faculty. This study has also attempted to test three aspects of student evaluations of faculty: sex, class level and black-white differences. The next section of the review will deal with specific literature relating to these three
Sex, Class Level and Black-White Differences

This section of the review will focus upon three important aspects of student evaluation. The three areas relate directly to the three major hypotheses and to the findings of this study. Research will be reviewed concerning sex, class level and black-white differences related to student evaluations.

Sex differences

Most of the data relating to male-female student differences in the evaluation of faculty has been a part of larger studies. The tendency has been for the researcher to add a dimension to his overall research package that deals directly with sex.

A study conducted by Elmore and LaPointe (1974) at Southern Illinois University involved a total of 1,474 courses conducted during the 1971 school year. Specific courses were matched on the basis of course number and sex of the instructor. From the total number of classes, 38 pairs of classes were made. The final pairing represents classes from a number of departments at the university. Using a two factor analysis of variance, the
researcher found no difference between male and female instructors, male and female students ratings or the interaction of the two. From the twenty item questionnaire used in the study, the researchers concluded about male-female student differences:

Only one significant difference between male and female students emerged. Female students rated instructors higher on Item 13, 'specified objectives of the course', than did male students (p. 387).

From the results of the study, the authors indicated that "neither of these differences, however, seems to be practically significant" (p. 387).

Touq and Feldhusen (1975) concluded from a study of eighteen instructors and 488 undergraduate students enrolled in eight different classes at Purdue University, that female students tended to rate instructors higher on three items of the Purdue Rating Scale For Instruction. The three areas significantly rated higher were: sense of proportion and humor, personal appearance and stimulating intellectual curiosity.

Bendig (1952), in a study carried out during the summer of 1951 at the University of Pittsburgh, found a significant difference between ratings of two instructors of three classes by the male and female students enrolled in the classes. Female students tended to rate both instructors more favorably than did male students. In
another study, Bendig (1953b) attempted to determine the relationship of two factors, instructional competence and instructor empathy, of the Purdue Rating Scale to student sex and course achievement. The researcher found a definite interaction between sex and instructor on Instructor Competence. Bendig failed to find any differences with the factor empathy. When both competence and empathy were combined, there were no significant differences in male and female ratings of instructors. The author concluded that "a particular instructor's competence may be rated quite differently by men and women students" (p. 432).

There are a number of other studies that have dealt with male-female differences in an indirect manner (Heilman and Armentrout 1936, Lovell and Haner 1955, Downie 1952, Carney and McKeachie 1966, and McKeachie, Lin and Mann 1971). However, these studies present similar results and therefore collectively fail to establish any definite pattern with regards to students sex and instructor ratings.

**Class level differences**

A student's class level has been one of the factors considered as having an effect upon student ratings of instructors. Many concerned with student ratings feel
there may be a distinct difference in instructor ratings based upon a maturity factor (Elliott 1949, Bendig 1952, Bryant 1967 and Lazovik 1972). Although the question hasn't been of primary concern, there are studies in which class level has been an added dimension.

In the study by Bendig (1952), a distinct difference between the combined ratings of freshmen and sophomores versus juniors and seniors was found. The study combined freshmen and sophomores into one group identified as lower academic level and juniors and seniors into the upper academic level and was limited to a comparison of those categories. An analysis of the mean scores for the two groups indicated that the upper academic students rated instructors less favorably than did lower academic level students.

The results of a study by Lazovik (1972) contradicted the findings of the Bendig study. The study by Lazovik was conducted at the University of Pittsburgh during the summer of 1971. A total of 144 Arts and Science faculty participated in the study. The problem of the study was to determine if students at different academic levels, freshman through seniors, differed in their judgement of their teachers. The author concluded that student ratings "at different levels of academic maturity (F, S, J, Sr.) indicates that this variable accounts for two-tenths of
one percent of the variance in student evaluations. It can be dismissed as of no consequence" (p. 2).

In yet another study, it was found that course level made a difference in instructor rating (Aleamoni and Graham, 1974). The purpose of the study was to determine if the tendency for faculty members of higher rank to receive higher ratings remained constant when class size and course level were considered. Data was collected from 488 courses with 488 distinct professors at the University of Illinois. The hypothesis that there was no difference in ratings by class level was rejected. Ratings made by freshmen, sophomores, juniors, seniors and graduate level classes made a difference in the instructor rating. In addition, an interaction between the size of a class and the level of the class was found to be significant.

Bryant (1967) in criticizing student evaluation of faculty noted that the class level can make a difference in faculty ratings. The author wrote:

A professor teaching a freshman course, filled with students taking it because they have to, is obviously at a disadvantage in competition with a professor teaching an advanced course to students who chose it as an elective (p. 328).

Finally, McKeachie (1969b) noted that Elliott in 1949 found that a student's undergraduate class level made little difference in his or her ratings of the college instructor. Graduate students on the other hand rated
their instructor higher than the undergraduate (p. 215).

From the review of literature regarding class level, it can be noted that the authors vary in their findings and conclusions. Therefore, the lack of consistency makes it difficult to determine if class level has an influence on student ratings of faculty.

Black-white differences

The third dimension of this study had to do with instructor differences as a result of the instructors skin color. As there were no studies that were directly concerned with student rating of black and white instructors, the review was limited to studies that related skin color and student evaluations.

Hutchison (1974) indicated that if educators were to use student ratings to improve education, they should understand the individuals who are directly involved. He stated that:

... any successful evaluation or improvement of teaching ought to develop after a careful consideration of the attitudes of those directly affected -- namely students and faculty members.

Since the college student and faculty live in a multi-cultured society, it seems most appropriate for educators to be aware of the attitudes that might influence student evaluations.

Katz et al. (1964) conducted a study to determine if
the race of the experimenter and the type of instructions given to a group of Negro boys would effect their hostility ratings as measured by a self-developed hostility scale. Seventy-two black students of high school and junior high age, 13-18 years old, volunteered for the experiment. The subjects were divided into 4 equal treatment groups. The results of the study indicated that the subjects reacted the same to the black and white experimenter and the type of instructions given. However, there was a significant interaction between the race of the experimenter and the instructions given. The researchers concluded that when the black experimenter gave test instructions rather than neutral instructions (no threat of test) the hostility scores of the blacks increased while with the white experimenter giving test instructions, the hostility decreased. The study indicated a change in behavior because of the skin color of the experimenter and type of instructions given.

In another study, Rokeach and Mezei (1966) ran three different experiments to determine if race or beliefs was more of a determinant for prejudice or discrimination. Two experiments were conducted on a university campus, using college students while the third involved job applicants for four different positions in two mental hospitals. The findings indicate:
(i) similarity of belief is a considerably more frequent basis of choice than disimilarity of belief; (ii) similarity of race is rarely a basis of choice - considerably less often even than chance and no more frequently than dissimilarity of race; and (iii) similarity of belief is a considerably more frequent basis of choice than similarity of race (p. 169).

The significance of this study was that the subjects discriminated against others because of different beliefs and not because of race. As a result one would expect students' ratings not to be effected by race but by other factors - one of which may be shared or common beliefs between the student and the faculty member.

Veldman and Peck (1969) designed a study to assess various aspects of pupil evaluation of student teachers. One of those aspects was the social class level of the school. Thirty-three schools in the Austin, Texas area were grouped into five socioeconomic levels according to the nature of the district. The student teachers were rated on five factors by their students. The results of the socio-economic element of the study revealed that student ratings of faculty differ on only one element, the factor of lively and interesting. The lower socioeconomic students rated their teachers higher on the lively and interesting factor than did high socioeconomic students.

In an earlier study conducted by Jackson and Fuller
(1966), at the University of Texas, social class was again one of the variables. They obtained social class origin information for each teacher included in the study. In addition, students were categorized into two social classes based on the neighborhood from which they were drawn. The results indicate that:

Lower-class teachers were evaluated more authoritarian by all pupils but particularly by lower-class pupils. In general, pupils seemed to prefer a teacher of a different social class except in evaluations reflecting effective communications, where teachers of the same social class as pupils were rated higher by their pupils (p. 2).

The two Texas studies indicate that there seems to be some difference in instructor ratings based upon social class.

Borland (1973) conducted a research project to determine if disadvantaged students admitted to institutions of higher education rated faculty differently. The subjects consisted of eleven students enrolled in a special summer program open to disadvantaged freshmen and ten freshmen enrolled on a regular basis. All students completed three evaluation instruments, each intended to measure one of the following: (a) instructor and course, (b) instructor characteristics and (c) classroom environment. In all three categories there were no significant differences between the ratings of the two groups of students. The author concludes that freshmen students
admitted into special programs actually responded in a similar manner as did other new freshmen.

"Does teacher's skin color make a difference" is the title of an article by Glick (1971). Glick sought to determine if the teacher's skin color made a difference in test performance for black youngsters located in an inter-city public school setting. There were a total of 4 treatments administered to 4 groups of eighteen randomly assigned eleventh graders. Using a two-way analysis of variance, Glick found no significant difference between any of the four groups.

Although there has been no studies done to determine the extent in which skin color effects student evaluations, the studies that were presented relate to student evaluation and skin color. Hostility, beliefs, social class, disadvantaged and test performance were all considered as factors being relating to student evaluations in a black-white context.
Student evaluations of faculty, hasn't been without criticism. In an attempt to overcome these criticisms, a number of educators have provided empirical data to judge the worth of these evaluations. A theoretical base for student ratings, the validity and reliability of such ratings and the usefulness of the evaluations will be reviewed.

**Theoretical base**

Although there has been a great deal written about student evaluation of faculty, there has been little related to theory. However a few studies have attempted to address the question of theory. These studies and studies that indicate a theoretical base will be discussed.

Although not always apparent or written, most studies are based upon theory. Voeks (1954) suggested that instructors differ in a number of observable ways. In a study of the top and bottom 10% of faculty rated by students at the University of Washington, Voeks found striking differences between the top and bottom category of teachers. These differences seem to indicate student ratings are based upon observable teacher behavior.
Gibbs (1955) prepared and administered to 119 male students a teacher behavior description scale intended to describe teacher behavior. The students rated seventy different male college teachers. As a result of a factorial analysis procedure, Gibbs identified 4 relatively independent factors or dimensions of teacher behavior. The 4 factors were: friendly democratic behavior; communication behavior; systematic, organization behavior; and academic emphasis. These 4 factors support the contention that college instructors ratings are based upon observable faculty behavior.

Although many authors noted that student evaluations were based upon a systematic rating of observed teacher behavior, others have found little theoretical value in this approach. Travers (1950) in reviewing two methods of measuring teacher effectiveness, indicated that student evaluation of faculty lack any solid theoretical foundation. The two methods for evaluating were: (1) the extent to which the teacher was able to develop behavioral objectives and later measure the extent to which the teacher was able to develop the desired behavior in the student and (2) ratings of the instructor based upon teacher behavior. In regard to rating faculty behavior as a criteria for teaching ability, Travers stated:
The procedure assumes that there is a known and well established relationship between behavior of the teacher and the development of the student (p. 41).

Travers further indicated that measuring teaching ability by means of observing teacher behavior was unjustifiable.

Vandervert (1974) developed a theoretical base for student evaluations. In his assessment of today's rating scales, Vandervert wrote about typical rating scale items:

These items even though they have become standards it must be admitted, have no well-founded theoretical basis related to the satisfaction of student needs other than an intuitive one (p. 1).

The author proposed a new model for student evaluation of faculty based upon student needs-to-be-satisfied. Vandervert indicated that most of the evaluation models attempted to address needs that were secondary. Using the established system, students were doing no more than rating faculty on items that indicated the student's satisfaction with the instructor. The rating failed to reflect the real reasons for the student being in the classroom. "Was the instructor well prepared for class meetings" may be an item that the students feel was important, but Vandervert would ask does it really satisfy the objectives for which the students were attending class. Vandervert's model is based upon three interrelated criteria:

(1) that the needs be related to the goals or objectives of instructors and the institutions which employ them, (2) that the satisfaction
of the needs be objectively measured on the instructor, and (3) that the needs be theoretically defendable in relation to needs college students in the college classroom actually do have (p. 3).

The author outlines the steps necessary to adopt the model for use in the college classroom.

**Reliability and validity**

The question of reliability and validity of student ratings are important items if one is to accept student ratings of faculty as a part of the educational process. Therefore, the reliability and validity of student ratings will be reviewed.

An article written by Costin, Greenough and Menges (1971) addressed the question of the reliability and validity of student ratings. The authors summarized most of the work completed before 1971. The article was used as a source for this review.

Remmers and Brandenburg (1927) in an effort to test the reliability of the Purdue Rating Scale devised two forms of the scale for use in the experiment. The students who participated in the study were asked to rate the instructor using one of the two forms, without their being informed of the second form. Three days later they were again asked to rate the instructor using the second form. The students were unanimous in their testimony that they
could not remember their first ratings. As a result of the experiment the two authors found that "student judgments as measured by the Purdue Rating Scale for Instructors had a considerable degree of reliability" (p. 523).

In three studies conducted by different researchers, similar results were produced. Drucker and Remmers (1951), Bryan (1963) and Centra (1973a) all tested the correlation between present student and alumni ratings of teaching. Drucker and Remmers found a correlation between .40 to .68 on specific items, while Centra reported a correlation of .75 and Bryan a .80 correlation.

Lovell and Haner (1955) obtained a correlation of .89 between ratings made by the same group of students two weeks apart. Costin (1968) found correlations of .70 to .87 on four dimensions of a student rating scale using mid-semester and end-of-semester ratings.

Unlike the reliability studies that have been reviewed, studies on validity are somewhat conflicting. However, several authors have addressed the question of validity.

In a study involving 18 instructors, 488 undergraduate students in 20 different classes, Touq and Feldhusen (1974) tested criterion-reference validity of
student ratings. The students' ratings were correlated with those of two trained experts. The researchers concluded that "the correlations found in the study indicate some agreement between students and expert observers with regard to instructor's classroom behavior" (p. 4). Therefore, there is some support for criterion-reference validity of student evaluations.

McKeachie, Lin and Mann (1971) analyzed the correlation between teacher ratings by students and six factors of the Introductory Psychology Criteria Test. The authors conclude:

"... our results ... do not invalidate the use of student ratings as one source of evidence about teacher effectiveness, but are less convincing than we had hoped for (p. 444).

McKeachie and Solomon (1958) attempted to validate student ratings of faculty by comparing student class ratings of a faculty member and the number of students in that class who later elected another class in the same discipline. Instructor ratings were significantly correlated with students electing to continue in the same field during two of the five semesters in which the study was conducted.

Miklich (1969) reported a study in which he was the teacher was evaluated by two different classes during the Spring Quarter, 1967 at the University of Hawaii. One of
the classes he was enthusiastic about teaching and the other he was disinterested in. Miklich attempted to test the validity of the Purdue Rating Scale. The experiment "allowed a test of students' ability to discriminate validly better prepared, more experienced, and more interested teaching" (p. 964). Miklich concluded that students can and do make valid ratings of teaching performance.

One of the criticisms of student ratings of faculty was that such ratings were directly related to student course grades. Steward and Malpass (1966) indicated in their analysis of 1,975 questionnaires, that students expecting high course grades rated their instructors significantly higher than did those expecting low grades. On the other hand, Langen (1966) reported just the opposite. From a survey of 30,000 evaluations, Langen found no relationship between student anticipated grades and teaching performance.

The question of the validity of student rating of faculty seems to be questionable. A number of authors have found several conflicting results. In addition, those studies that dealt directly with the validity of student ratings did little to confirm the suspicion that student ratings are valid measures of teacher effectiveness. The reliability studies on the other hand seemed
to indicate a consistency in how students evaluated faculty. As a final note, McKeachie (1970) offered the following:

I believe student evaluation of teaching can be valid and useful; but let us remember that the ultimate purpose of evaluating teaching is to improve learning. Evaluation is not an end in itself. If a program of evaluation creates anxiety that interferes with good teaching, if it stimulates or reinforces hostility, if it simply takes so much time from learning that the net gain is negative, let's forget it. We must weigh the cost of evaluation against the gains. I believe there can be important gains but I would not overlook cost. The college is a learning community. Evaluation of either students or teachers should be forced to justify its existence in terms of learning (p. 10).

Use of evaluations

Probably the most important phase of student evaluations is the feedback or use of evaluations phase. How the evaluations are used, what they are used for and if they really make a difference in teacher behavior are aspects that must be addressed.

McKeachie (1969a) listed several uses of student evaluations, each dependent upon the goal one has in mind.

1. Administrative purposes
2. Improve teaching
3. Promote student thinking about education
4. Assist students choice of course and instructor
Smock and Crooks (1973), in contrast to McKeachie indicate
somewhat different uses of student evaluations:

1. Instructor use
2. Student use
3. Department heads use
4. College and campus administrators use

Centra (1972a) in relation to McKeachie and Smock and
Crooks, suggest 2 main reasons for evaluating teaching:

1. To help make decisions about whom to promote
2. To improve instruction

The McKeachie, Smock and Crooks and Centra studies all
detail a comprehensive system for the use of student
evaluations.

In a survey conducted by Costin, Greenough and Menges
(1971) of 404 students enrolled in psychology courses at
the University of Illinois, it was found that most students
feel that student ratings are valuable. In addition,
these same students indicated that they agreed somewhat
that student evaluations will effect future teaching
performance.

Centra (1972c) conducted a study to determine if
faculty ratings of themselves were much different than
their students. He suggests if there was such a dif­
ference then there was a need for instructors to be told
how their students rate them. Centra found little agreement between instructor self evaluations and their students' evaluation. The median correlation for 17 items was .21. Therefore, the "discrepancies between the 2 sets of ratings . . . underscore the need for student feedback . . ." (p. 1).

Whetstone (1974) developed a system of student feedback for use at the University of Colorado. The author writes:

... the chief advantage of this procedure is that each instructor of a course receives detailed and comprehensive information that is individually specific and at the same time allows comparisons with various norm groups within the University. The instructor also receives subjective responses by students on the backside of the questionnaire which asks for the most and least effective aspects of the course and utilizes the critical incidence technique (p. 1).

The information the instructor receives is a print-out of his or her performance along with normative data used for making comparisons.

A feedback system of student evaluations is useless unless the classroom teacher uses the feedback information. Four research reports have attempted to determine the use of student evaluation in changing teacher behavior.

Centra (1972b) studied the difference between three groups of teachers from five diverse colleges. One group received feedback, a second no feedback and the third was
posttest only. The students in the feedback and no feedback groups completed mid-semester evaluations for their instructors. The instructors of the feedback group received a report of the student midterm evaluation. An end of term evaluation was given to all three groups. The final evaluation indicated no significant difference between the three groups.

Feedback, in another study (Vogt and Lasher, 1973) also found that student evaluations did not improve teaching. The two Bowling Green researchers compared instructor ratings over a period of time. They contended that if student evaluation had improved teaching then there should be an increase in instructor rating over a period of time. These samples showed no increase.

In another study (Centra, 1973a) of 400 faculty members, it was found that student ratings did have some effect upon instruction. Those instructors who were identified as unrealistic in how they viewed their teaching, changed after a half semester. Other instructors also changed, but over a longer period of time.

Yet in another study (Lazovik, 1975) feedback was shown to be valuable in improving instruction. Using decile standings for 50 teachers who had taught the same class at least a term apart showed that student evaluations did improve teaching. Lazovik wrote:
The findings here indicate that when student evaluations of college teaching are provided for those faculty who want to improve, higher evaluations result for a significant number of teachers who continue to request that their students evaluate them (p. 35).

Lazovik's study was conducted at the University of Pittsburgh.

The authors are in conflict as to the usefulness of student ratings. Although this conflict exists, it was noted that evaluations are being used by faculty, administrators, and students for self improvement, promotion, and class selection.

Blacks in Higher Education

The hiring of black faculty members is a relatively new occurrence in higher education. One of the reasons for this basic change is government pressure.

Federal legislation

The 1964 Civil Rights Act was the first real move against segregation in public education (Title IV). The law covered all public schools, including all institutions of higher education. The authors of the law stated in Section 401c:

... public college' means any institution of higher education or any technical or vocational school above the secondary school level, provided that such public school or public college is operated by a State, subdivision of a State, or
governmental agency within a State, or operated wholly or predominantly from or through the use of governmental funds or property, or funds or property derived from a governmental source (p. 247).

The 1964 Civil Rights Act led the way for equal opportunity in higher education.

In addition to the 1964 law, former president Lyndon B. Johnson signed Executive Order 11246 entitled "Equal Employment Opportunity (1965) to combat discrimination on the basis of race, creed, color or national origin. The law not only detailed guidelines for nondiscrimination in government employment, but nondiscrimination in employment by government contractors or subcontractors. College campuses by virtue of being a federal contractor has had to comply with the Civil Rights law and Executive Order 11246.

In 1972 the Civil Rights Act was amended. Title VII of the act outlawed discrimination by any employer of fifteen or more persons and also allowed aggrieved individuals and groups to take a case to court. These two rulings according to Ducey (1974) "are of such vigor and scope that their impact on colleges and universities is likely to be much greater than that of HEW efforts" (p. 2). The enforcement of the 1964 law as it related to education was the responsibility of the Department of Health Education and Welfare until the 1972 amendment.
The Equal Employment Opportunity Commission was given jurisdiction over educational agencies in 1972. One result of the above legislation has been an increase in the number of black faculty in higher education. In order that the law be carried out, an affirmative action program was established.

**Affirmative Action**

Affirmative action was the result of the concerns of many Americans. These Americans saw Affirmative Action as a means of assuring equal rights for all citizens. Hannah (1966), Chairman of the 1966 United States Commission on Civil Rights and President of Michigan State University wrote:

> Is there really any question whether the public universities will participate in working out the solutions to this, the greatest domestic problem of our times? A realistic appraisal of the situation leads to the conclusion that the public universities have no choice; that they must enlist - or be drafted - to serve this cause of social improvement, as they have served so many others in the past (p. 62).

To ensure that many institutions moved "to serve this cause of social improvement", the federal government made Affirmative Action the concern of all institutions who were prime contractors or subcontractors of the federal government.

Part 60-2, Title 41 of the Code of Federal Regula-
tions outlined the Affirmative Action Program. The pro-
gram was divided into four parts. Upon adopting an Affirma-
tive Action plan, the contractor had to establish goals
and timetables for accomplishing the objectives. Code
60-2.12 Section (a) (1975) stated:

The goals and timetables developed by the con-
tractor should be attainable in terms of the
contractor's analysis of his deficiencies and
his entire affirmative action program. Thus,
in establishing the size of his goals and the
length of his timetables, the contractor should
consider the results which could reasonably be
expected from his putting forth every good faith
effort to make his overall affirmative action
program work (p. 221).

Those contractors affected by Affirmative Action not only
must establish a program but must make the program work.

There has been a number of articles and booklets
written regarding Affirmative Action. These booklets
explain the law and usually delineate suggestion for
establishing and maintaining an Affirmative Action program.
One such booklet published by the National Education
Association (1973) list five major items that are usually
included as a part of an Affirmative Action policy:

1. Purpose or intent
2. Analysis of work force utilization
3. Problems and methods for solution
4. Carrying out the plan
5. Establishing numerical goals and timetables
Affirmative Action programs are a part of the colleges responsibility. Affirmative Action has also effected the hiring policies used by large universities. Therefore, many universities have made commitments to Affirmative Action.

Chapter Summary

The review of the literature has centered around four areas - student evaluations in higher education; sex, class level, and black-white differences; theory, reliability, validity and use of student evaluations; and blacks in higher education. Typical studies in student evaluations of faculty were presented. The conflicting results of studies dealing with students sex and class level as factors in instructor ratings were reviewed. Studies relating skin color and evaluations were summarized. The controversy over a theoretical base for student evaluations, the conflicting findings relating to validity of student ratings, reports dealing with student rating reliability along with the lack of consensus regarding the use of student evaluations were discussed. Finally, the chapter dealt with federal legislation and Affirmative Action programs responsible for the increase number of black faculty in higher education.
CHAPTER III. RESEARCH DESIGN

The problem of the study was to ascertain if skin color made a difference in instructor ratings by predominantly white classes in a predominantly white university. This chapter describes the methods and procedures used to resolve that problem. The chapter has been divided into three major parts: (1) the instructor rating form; (2) video tape production; and (3) treatment. Each major heading is subdivided for clarity. This chapter should serve as the basis for accepting the results presented in Chapter IV.

The Instructor Rating Form

As indicated in Chapter I, it was important to gather information for which to test the hypotheses. In addition, the method used to gather the information had to simulate the instructor rating techniques used on most college campuses. The first step in the research design was to build an instructor rating form. Because of the uniqueness of the research design, existing scales were inappropriate. Wilbert J. McKeachie (1969b) indicated:

... courses differ from one another, questions of particular interest to one instructor are irrelevant to the aims of another. Consequently many instructors may wish to construct their own scale (p. 219).
Considering the unique aspect of using video tapes presentations as the treatment, a new instructor rating scale was developed. The procedure followed was to review selected rating forms and rating scales in order to develop an instructor rating form for this study.

Review existing forms

In an effort to develop an instructor rating form that would be appropriate, a number of university instructor rating forms were reviewed. The rating forms selected for review were based upon correspondence (Appendix A) with Dr. Wilbert J. McKeachie, Department of Psychology, The University of Michigan. Dr. McKeachie has published extensively in the area of student evaluations of faculty. Dr. McKeachie indicated five people who had been most influential in developing material related to student evaluation of faculty (Appendix B). These five people were contacted by mail for information and normative data on scales that they had or had helped develop. All five responded to the request and sent existing scales as well as data on each scale. Two of the five individuals had worked on one of the scales received. As a result, a total of four scales were reviewed on the basis of Dr. McKeachie's suggestions. The four scales were:
1. Purdue Scale
2. University of Minnesota Scale
3. University of Pittsburgh Scale
4. The University of Illinois Scale

The Purdue Scale is better known as the **Instructor and Course Appraisal Cafeteria System**. The Cafeteria System used at Purdue is unique in that it allows the individual instructor to select up to 40 items from which to be evaluated. The total number of items the instructor may select from is 200. In addition to the 40 items an instructor may select, the course and instructor are also evaluated by the students on five "university core items". Departments may also develop a "department core" using the 200 items from the Cafeteria System.

The major areas of the Cafeteria System are:

1. Clarity and effectiveness
2. Student interest/involvement in learning
3. Broadening student outlook
4. Teaching/learning of relationships and concepts
5. Instructor provides help as needed
6. Providing feedback to students
7. Adapting to individual differences
8. Respect and rapport
9. Course goals or objectives
10. Usefulness/relevance of content
In addition to the 20 broad headings, the specific university core items are:

1. My instructor motivates me to do my best work.
2. My instructor explains difficult material clearly.
3. Course assignments are interesting and stimulating.
4. Overall, this course is among the best I have ever taken.
5. Overall, this instructor is among the best teachers I have known.

The Purdue Cafeteria Rating Systems allows for a lot of flexibility. The instructor has a wide range of statements from which to select.

The University of Minnesota has four student opinion
surveys (rating forms) to assess a number of aspects relating to student evaluations. The four are:

1. Student Opinion Survey - General
2. Specific Student Opinion Survey - Reading Material
3. Specific Student Opinion Survey - Test and Grading
4. Student Opinion Survey - Supplement

Of the four evaluation instruments, the Student Opinion Survey - General was most often used. The statements and questions on the form that deal specifically with instructor evaluation were:

1. Clearly presented the subject matter
2. Was approachable
3. Got me interested in her/his subject
4. Raised challenging questions
5. When appropriate, related course material to other areas of knowledge
6. How much did you like the instructor as a person?
7. How would you say you learned from this instructor?

The Center for the Evaluation of Teaching at the University of Pittsburgh has developed a Student Opinion of Teaching Questionnaire which allows instructors to add up to ten items of his or her own formulation. The basic questionnaire contain a standard ten questions used
by all instructors to measure instructor behavior. The ten items were:

1. Interprets difficult or abstract ideas clearly
2. Makes good use of examples and illustrations to clarify concepts
3. Conveys his/her knowledge of the subject to students
4. Includes in class worthwhile and informative material which is not duplicated in the text
5. Has presented course content so that I perceived its relevance to my interest
6. Has presented course content in an organized manner
7. Has increased my interest in the subject
8. Has given me new viewpoints or appreciations
9. Has stimulated thinking on the part of students
10. Using the same five categories above, mark the number which best express your judgment of the instructor's overall teaching effectiveness as compared with other instructors you have known.

The University of Illinois Course Evaluation Questionnaire contains 23 statements that relate to the course or instructor. As only the instructor evaluation aspect of
the questionnaire was of interest, only items relating to the instructor were reviewed. Those statements concerning the instructor were:

1. I would take another course that was taught this way
2. I would have preferred another method of teaching in this class
3. It was easy to remain attentive
4. The instructor did not synthesize, integrate or summarize effectively
5. The instructor encouraged the development of new viewpoints and appreciations
6. I learn more when other teaching methods are used
7. The instructor was excellent
8. The instructor demonstrated a thorough knowledge of the subject matter
9. I would rather not take another course from this instructor
10. Some things were not explained very well
11. The instructor seemed to consider teaching as a chore or routine activity

In addition to these specific questions, students are also asked to make written comments on a number of other items. One of those items related to the student's perception of his instructor. The question asks "what are
your general comments about the instructor(s) in this course?"

In addition to the four scales above, the Test Analysis And Development Corporation in Boulder, Colorado was contacted for information on the Course Evaluation Questionnaire. There were three parts to the Questionnaire; part one dealt with the instructor, part two the course, and the third part was for supplementary questions. Part One (the instructor) was of most concern. The instructor aspect contained 24 statements that described teacher behavior. An example of the items are:

1. Discusses points of view other than his own
2. Identifies what he/she considers important
3. Has a sense of humor
4. Respects students as persons
5. Explains clearly
6. Invites criticism of his/her own ideas
7. Is a dynamic and energetic person
8. Is enthusiastic about his/her subject
9. Presents origins of ideas and concepts
10. Has an interesting style of presentation

In addition to the twenty statements describing instructor behavior, an additional eight may be added in Part Three (supplementary questions) of the questionnaire.
In addition to the above forms, the instructor evaluation form used at Iowa State University was reviewed. Because of the familiarity of the form to the students at Iowa State (the population from which the sample used in the study was drawn), the construction and layout of the Iowa State form was of particular interest. The instructor rating form used at Iowa State is centered around six areas. Those six areas are:

1. Organization/Efficiency
2. Attitude
3. Student Interest
4. Interaction
5. Explanation
6. Evaluation

The Iowa State rating form is located in Appendix C.

In total, six of the most outstanding student evaluation of faculty forms from across the country were reviewed in order to build a form for use in this research project.

Rating scale

One of the important factors in building a rating form that could be used in this study was to select an appropriate rating scale. Although many of the scales on the forms reviewed differed on the range and value
assigned each response, they all used a Likert-Type Scale. The University of Illinois used a four item letter scale which ranged from strongly agree to strongly disagree. The University of Minnesota used a number scale ranging from 1 to 7. The Test Analysis and Development Corporation developed a five item number scale which ranged from 0 to 4. The University of Pittsburgh, Purdue University and Iowa State University also used a five-point scale except the range was from one to five. Although the Pittsburgh, Purdue and Iowa State Scales used different words to describe the meaning of each number, a mark of five was considered the highest possible rating.

From the review of rating scales used on typical instructor rating forms, a Likert-Type Scale was developed. The range of scores and the interpretations of the number values were difficult to determine. Because of the familiarity of the Iowa State Scale to the subjects that would be used in the study, the range and value distinctions as outlined on the Iowa State Scale were adopted.

Development of the rating form

Using the six forms reviewed as a basis for developing a new form, a total of thirty-one rating items were first written. This first form served as a base from
which the final form was eventually developed.

From a discussion with the director of the Testing and Evaluation Office at Iowa State University, and other university personnel, it was apparent that a less demanding form needed to be developed. A number of items were eliminated from the first form and several others were combined into one general statement describing the instructors (lecturers) behavior. As a result, the form was reduced to a total of twelve items that seemed most appropriate to the study.

In addition, four identification items were added for administration and analysis purposes. These four items were completed by the students before they evaluated or rated the instructor. A copy of the form is located in Appendix D.

Rating instructions were written in order to clarify the procedure used to rate the instructors. The first four instructor rating items on the form were basically the same as the Iowa State University form. The only exceptions were the words instructor and course were replaced with lecturer and lecture on the rating form developed for this study. Items 10 through 16 were not related to the Iowa State form, but were the result of similarities between the rating forms reviewed. The areas
evaluated were:

1. Organization/efficiency
2. Attitude
3. Student interest
4. Interaction
5. Explanation
6. Lecture
7. Visuals
8. Presentation
9. Knowledge
10. Lecturer
11. Class acceptance
12. Individual acceptance

The instructor rating form developed was a result of first reviewing existing forms, developing an appropriate rating scale and organizing and developing a workable form for use in this study. The process used to develop the instructor rating form resulted in the establishment of content validity.

Video Tape Production

The second stage of this study dealt with the development of the video tapes to be used as the treatment. Tapes had to be developed that would yield accurate
data. As was stated earlier, an attempt was made to equalize all relevant variables except for skin color. As a result, the selection of lecturers (instructors) and script, methods used in taping, selection of experts to evaluate tapes, methods used to review tapes, and analysis of the experts scores were considered as important phases of the video tape production.

Selection of lecturers (instructors)

The selection of instructors was the first phase in developing equal tape presentations. Criteria for the instructor selection was written. The criteria were written based upon variables that had to be controlled in order to equalize instructor differences. The criterion used with the most important criteria listed first were:

1. Skin color (typical of each ethnic group)
2. Age
3. Mannerisms (audio and physical)
4. Physical appearance
5. Teaching experience
6. Educational philosophy
7. Willingness to work
8. Availability

Two individuals who met or were close on all criteria
listed were selected. Items 1, 3, 6, 7 and 8 were subjective evaluations and therefore had to be determined by an interview. Items 2, 4 and 5 were objective and could easily be compared. The instructors selected and their objective qualifications (Items 2, 4 and 5) were:

<table>
<thead>
<tr>
<th>White Instructor</th>
<th>Black Instructor</th>
</tr>
</thead>
<tbody>
<tr>
<td>23 years old</td>
<td>26 years old</td>
</tr>
<tr>
<td>6'1&quot;, 175 lbs.</td>
<td>6'0&quot;, 200 lbs.</td>
</tr>
<tr>
<td>1 year teaching experience</td>
<td>1 1/2 years teaching experience</td>
</tr>
</tbody>
</table>

**Selection of script**

Although there were no rigid restrictions as to the content of the scripts, the content had to be viewed by the instructional staff at the university as a meaningful use of time. As the tapes were shown to a number of different types of classes that varied in subject matter, age, sex, grade level and interest, the subject of the lectures also had to be of a general nature.

A number of topics were considered for the scripts. Metrics, sex education and drugs were topics that were considered. Metrics was thought to be lacking in sufficient interest and sex education as a topic may have caused students to rate whites and blacks differently. Drugs seemed to be most appropriate. The content of the scripts was taken from a government publication.
entitled *Drugs In Our Schools* published in 1973. The subject matter was thought to be of interest to all students. Whether one was a graduate student and parent or a new college freshman, drugs were a part of his or her social environment. The script was written in two parts, the major headings as outlined in each part were:

**Part I**
1. The scope of the (drug) problem
2. Drugs abuse surveys
3. National drug arrests
4. The American family tragedy

**Part II**
1. The drugs used by the high school students
2. Programs for youth drug users
3. The school's response

Each instructor had to be familiar with the information in each script. The scripts were written and revised a number of times before and after the first two taping sessions. A copy of the final script is located in Appendix E.
Methods used in taping

In addition to producing tapes that were equal in presentation, it was necessary to try and produce tapes that were interesting. Therefore, camera and instructor ques were written into the scripts. As two cameras were used, the instructors movements were synchronized with that of the camera operators. In addition, closeups and fadeouts were included as part of the production.

To produce the finished tapes, over 15 takes were made during the three separate taping sessions. Both instructors were limited in their freedom of physical movement in order that physical movement would not distract or influence the student's ratings. In addition, the instructors were not allowed to deviate from the written script. This was done to ensure similarity in method and type of presentation.

Selection of experts

The experts for the tape evaluation phase were selected on the basis of a number of different items. The criteria used were:

1. Knowledge or awareness of typical student behavior
2. Some experience with student ratings
3. An earned doctorate in a field related to student personnel work

In addition to these three requirements, a fourth group requirement was that one of the three experts had to be black. All three requirements along with the group requirement were met. The first expert was the director of Test and Evaluation Service on the campus. The second expert was a professor of counseling in the Department of Professional Studies at the university. The third member was a professional counselor in the Student Affairs Division at Iowa State. All three members had had some experience and involvement in student ratings of faculty.

Methods used to review tapes

A total of two meetings were held with the experts for the purpose of reviewing the video tapes. After the first filming session, the tapes were evaluated by the experts. After each part was shown, the experts were asked to make independent judgments of the instructors based upon the student rating scale developed. After parts one and two of both tapes were rated, comments were made as to the quality of the instruction. Some of the comments made by the experts were: too much eye contact with the script, both instructors were dull in their presentations, and the black instructor's pace was much
quicker than the white instructor's. Although both instructors were close in terms of their presentation, they both needed to improve their overall presentations. In addition, the presentations were a little long. Adhering to the suggestions made at the first evaluation session, the tapes were again filmed.

The second session with the experts followed much the same procedure as the first. Upon completion of the second evaluation of all four parts, the experts verbally agreed that both instructors were equal in their presentation.

Analysis of expert scores

Although the three experts chosen to review the tapes for this study all verbally agreed that the final tape presentations met the requirements necessary to be used in the experiment, a statistical analysis was performed. The experts made written ratings using the same scale and instructor rating form the sample of the study would use.

The statistical analysis of the experts scores was difficult. As there were only three experts used in the evaluation process an N of three made it most difficult to draw meaningful conclusions. However, the mean scores of the experts on all four parts of the presentation were
placed in a $3 \times 4$ randomized block design with one replicate per cell (Dayton 1970, p. 163). Table 3 illustrates the design and also indicates the experts' means scores for each presentation.

Table 3. Mean scores of each expert for each tape presentation

<table>
<thead>
<tr>
<th>Presentations (Treatment)</th>
<th>White ($W_1$)</th>
<th>Black ($B_2$)</th>
<th>Black ($B_1$)</th>
<th>White ($W_2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tape 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experts 1</td>
<td>3.2</td>
<td>3.0</td>
<td>2.9</td>
<td>2.8</td>
</tr>
<tr>
<td>2</td>
<td>3.1</td>
<td>3.0</td>
<td>2.9</td>
<td>3.0</td>
</tr>
<tr>
<td>(Blocks) 3</td>
<td>3.3</td>
<td>3.2</td>
<td>3.3</td>
<td>3.2</td>
</tr>
</tbody>
</table>

The experts' mean scores on the instructor rating scale were used in the analysis. The results of the analysis is given in Table 4.

Table 4. Analysis of variance summary table for experts ratings

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>3</td>
<td>.0867</td>
<td>.0289</td>
<td>3.57</td>
</tr>
<tr>
<td>Blocks</td>
<td>2</td>
<td>.2950</td>
<td>.1225</td>
<td></td>
</tr>
<tr>
<td>Interaction</td>
<td>6</td>
<td>.0483</td>
<td>.0081</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>14</td>
<td>.3800</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
An F of 3.57 is nonsignificant at the .05 level. Therefore, there were no significant differences between the four treatment dimensions. Both parts of each tape were equated as equal.

The assumption necessary to complete the above analysis according to Dayton (1970) has been met:

Because of the lack of any within-cell variability, there is no estimate of an error means square and no separate within-cell-error terms to pool. Thus, it becomes irrelevant to test for homogeneity (p. 165).

The analysis of the experts scores indicate that the tapes used in the study appear to be equal using the instruction rating form developed for this study. As a result, the tapes were deemed worthy for use in this experiment.

Treatment

The next step was to administer the treatment. This involved selecting a sample within a given population, running a pilot study, contacting the instructors of the classes selected, administering the treatment, collecting the data and analyzing that data.

Description of the population

The population selected for use in this study had to represent a typical predominantly white college or
university. Iowa State University, the institution identified as being a predominantly white university, had a student enrollment of 19,464 students for the Spring Quarter 1976. Two hundred and thirty-six of these students were classified as American black minority students. The black minority population represented about one per cent of the total student body. The population was identified as the total number of university classes listed in the Schedule of Classes for the Spring Quarter 1976 at Iowa State University.

Selection of the sample

Using the Spring class listing from the Schedule of Classes, the population was stratified on the bases of class level and type of program. The class level strata was:

1. Graduate classes
2. Upper division classes
3. Lower division classes

The program strata included:

1. Physical sciences
2. Social sciences

The second strata was necessary in order that two classes could be chosen from each category. It was felt that this second stratification would assure classes being
selected from both the physical and social sciences. The classification of classes into physical and social science groups was difficult and therefore criteria had to be written for categorizing the classes.

<table>
<thead>
<tr>
<th>Physical Science</th>
<th>Social Science</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Thing originated</td>
<td>1. People originated</td>
</tr>
<tr>
<td>2. Deals with laws and principles of the physical world</td>
<td>2. Deals with social, organizational systems and aspects of human life</td>
</tr>
<tr>
<td>3. Deals with people indirectly</td>
<td>3. Deals with things indirectly</td>
</tr>
</tbody>
</table>

This criteria made it easy to stratify university classes on the basis of physical and social sciences.

In total, six stratified groups were formed from which to select the sample. The six were:

1. Graduate classes - Physical Sciences
2. Upper division classes - Physical Sciences
3. Lower division classes - Physical Sciences
4. Graduate classes - Social Sciences
5. Upper division - Social Sciences
6. Lower division - Social Sciences

A total of two classes from each group was needed for administration of the treatment.

One aspect of the design dealt with the difference between male and female ratings of instructors. Therefore, classes selected had to have at least a twenty-
eighty ratio of males to females or females to males enrolled. This couldn't be determined until the instructors of the classes chosen had been contacted.

A total of eight classes were chosen from each of the six groups. Eight were selected with the understanding that some classes wouldn't meet the twenty-eighty ratio and that some may not want to or could not participate in the study. The total number of classes selected in random order was forty-eight.

**Pilot study**

In order that the techniques for conducting the research in the classrooms be void of procedural error, a pilot study was conducted. Walter R. Borg and Meredith D. Gail (1963) support the pilot study:

> It (the pilot study) provides the research worker with ideas, approaches, and clues not foreseen prior to the pilot study. Such ideas and clues greatly increase the changes of obtaining clear-cut findings in the main study (p. 61).

An instructor of a university class that wasn't selected in the original sample to receive the treatment was asked to participate in the pilot study. The procedure and techniques that would be used in the classroom were tested in the pilot study. From the pilot study a few procedural changes were made. It was found that instructor rating answer sheets should be returned to the researcher after
each individual instructor was rated and that the re-
searchers reading of the instructions printed on the 
rating form wasn't favorable. In addition, it was found 
that students should be reminded that their judgment were 
all that was necessary. A few students during the pilot 
study were conferring with other classmates as to 
ratings given each instructor.

Instructor contacts

As none of the classes selected were required to 
participate in the study, it was important to outline the 
nature of the study and explain the worthiness of the re-
search to each instructor whose class was selected. In 
an effort to encourage each instructor's support, a letter 
of introduction and encouragement from the vice president 
for research, Dr. D. J. Zaffarano, at Iowa State Uni-
versity, was obtained (Appendix F). The letter from Dr. 
Zaffarano was used to introduce the research being con-
ducted.

Each instructor was contacted via the phone and 
informed of the nature of the research project. They 
were also informed that their class was one of several 
randomly selected for use in the study, and the total 
class time needed to conduct the study was approximately 
fourty minutes. Once the instructor agreed to having his/her
class participate in the study, the instructor was asked if their class could meet the 20-80 ratio. If the class met the requirements, a time was scheduled for the administration of the treatment.

Just as some of the classes selected in the random drawing could not meet the 20-80 sex ratio, a few declined to participate on the basis the instructor couldn't spare the extra class time. As a result, a total of eighteen classes were passed over. The procedure was to contact all classes in order of the random selection until two classes from each of the six groups could be selected.

**Administer treatment**

All the equipment necessary to administer the treatment was taken into the instructor's classroom. Therefore, instructors and students were not required to leave their typical classroom environment. The students were told that the reason for the evaluations of the two instructors (lecturers) was to help a research team determine if the two instructors were sufficient in their presentations to be included in a future government research project. The subjects were also told that this was an important phase of the research project and therefore, a number of classes were being asked to rate the two
instructors. Each class was cautioned to evaluate only the instructor and not the quality of the tapes or the content of the lecture.

As indicated in Chapter I, there were two different tapes with the order of presentation reversed on each tape. The scheme was to show two of the four classes in each class level tape one, and the other two classes tape two. The effect of seeing either tape one or two first would equal out over the four classes.

The procedure was to show the first part of either tape one or tape two (depending on the class) and then have the students rate the instructor using the instructor rating form. Each participant was asked to read the instructions to himself or herself before rating the instructors. After rating part one, the first answer sheets were collected. Part two was then shown to the total group. At the end of part two, a new answer sheet was distributed and the students were asked to rate part two of the lecture. At the completion of the second rating, the second group of answer sheets were collected along with the instructor rating forms.

After collecting all the data for a particular class, it was coded as to instructor, tape part number, level of class and separated by male and female responses.
Analysis of data

An analysis of the data first included a test of the assumptions necessary to meet the requirements for using the analysis-of-variance procedure for repeated measure designs.

Once the assumptions were met for a repeated measure design, the next step was to proceed with the analysis of the data collected. All individuals in each class filled out a rating form for both the black and white instructor. The individual scores for the white instructor were combined and those for the black instructor were also combined to determine the classroom mean for each instructor. This procedure was accomplished with the aid of the Test And Evaluation Service at Iowa State University.

Chapter Summary

Chapter III delineated the detailed information necessary for a complete understanding of the research study. The instructor rating form, how it was developed, along with the scale used was reviewed. The selection of instructors for the video taping production, the script, and method used to produce the tapes (treatment) were important elements in the design of the research. The selection of qualified experts, the experts' rating of the
tapes and the analysis of the experts' scores were areas also covered. The treatment included identifying a sample within a given population, running a pilot study, contacting instructors of classes randomly selected, administering the treatment and statistically testing the results. This chapter should serve as a basis for accepting or rejecting the findings outlined in the next chapter.
CHAPTER IV. FINDINGS

The findings of this chapter will be reviewed in light of the hypotheses presented in Chapter I. As indicated in Chapter I, a factorial design with one repeated measure and an analysis-of-variance procedure were used to test the hypotheses. Basic background data will be presented concerning the classes utilized, and the test of assumptions relevant to the analysis. The test of the three main effects (instructor, sex and class level) along with the interaction analysis will also be presented.

Classes Utilized

From a population of all Spring Quarter 1976 classes at Iowa State University, a total of 12 randomly selected classes were utilized. The sample represented classes from both the physical and social science areas. In addition, each class selected had to have at least a 20-80 ratio of males to females or females to males. Table 5 indicates the number of classes selected from each science area and the sex ratio within each class. The total number of individuals rating the black and white instructors in each class is represented by the numbers in the total columns of Table 5.
Table 5. Total number of males and females in each class rating instructors per level and science area

<table>
<thead>
<tr>
<th>Level</th>
<th>Physical Science</th>
<th></th>
<th>Social Science</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
<td>Total per class</td>
<td>Male</td>
</tr>
<tr>
<td>Lower Division</td>
<td>4</td>
<td>5</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>Upper Division</td>
<td>11</td>
<td>4</td>
<td>15</td>
<td>20</td>
</tr>
<tr>
<td>Graduate Division</td>
<td>16</td>
<td>15</td>
<td>31</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>3</td>
<td>8</td>
<td>4</td>
</tr>
</tbody>
</table>

Each individual in each class completed an instructor rating form. All male scores were combined together and all female scores were combined in each class to obtain a male and female mean score for each instructor in each class. This was necessary because classrooms and not individuals were randomly selected. According to Dayton (1970):

... the individual students in an intact classroom cannot be considered independent experimental subjects since they did not arrive in the classroom by a random process. Thus, the mean of the classroom scores should be used (p. 117).

The male and female mean scores in each class for both the black and white instructors are identified in Table 6.
Table 6. Mean scores for each class used in the study

<table>
<thead>
<tr>
<th>Instructor</th>
<th>Division</th>
<th>Mean Score</th>
<th>Mean Score</th>
<th>Mean Score</th>
<th>Mean Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower</td>
<td>Male</td>
<td>Female</td>
<td>Upper</td>
<td>Male</td>
</tr>
<tr>
<td>Black</td>
<td>Division</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructor</td>
<td>1</td>
<td>3.202</td>
<td>3.167</td>
<td>2.638</td>
<td>3.060</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2.798</td>
<td>2.910</td>
<td>2.963</td>
<td>3.042</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2.932</td>
<td>2.954</td>
<td>2.700</td>
<td>2.736</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>3.083</td>
<td>2.874</td>
<td>3.125</td>
<td>2.792</td>
</tr>
<tr>
<td>White</td>
<td>Instructor</td>
<td>2.604</td>
<td>2.600</td>
<td>2.578</td>
<td>2.929</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2.833</td>
<td>2.575</td>
<td>2.657</td>
<td>3.118</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2.334</td>
<td>2.750</td>
<td>2.158</td>
<td>2.118</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2.600</td>
<td>2.356</td>
<td>2.863</td>
<td>2.563</td>
</tr>
</tbody>
</table>

As a repeated measure was used on the instructor dimension, mean scores for classes 1 through 4 in each division can be compared.

Test of Assumptions

In order that the data collected could be analyzed, the homogeneity of variance and the homogeneity of covariance assumptions had to be met. Using the discriminant analysis procedure as outlined in *A User's Guide To The Statistical Analysis System* (1972), a variance-covariance matrix was developed to test the homogeneity of variance and covariance. The result of the test was nonsignificant with Chi-square of 13.381 and 15 degrees of freedom. As a result, the assumptions of homogeneity of variance and covariance were met.
Analysis of Data

The statistical testing of the hypothesis was accomplished using the ANOVA procedure as outlined in the SAS Manual (1972). The results of the procedure is given in Table 7. This will be the basis for which the discussion of the findings is related.

Table 7. The analysis of variance summary table for black and white instructor ratings

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among Subjects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class (A)</td>
<td>2</td>
<td>.155</td>
<td>.078</td>
<td>.750</td>
</tr>
<tr>
<td>Sex (B)</td>
<td>1</td>
<td>.032</td>
<td>.032</td>
<td>.311</td>
</tr>
<tr>
<td>AB Interaction</td>
<td>2</td>
<td>.036</td>
<td>.018</td>
<td>.173</td>
</tr>
<tr>
<td>Subjects (S)</td>
<td>18</td>
<td>1.863</td>
<td>.104</td>
<td></td>
</tr>
<tr>
<td>Within Subjects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructor (C)</td>
<td>1</td>
<td>1.500</td>
<td>1.500</td>
<td>54.279</td>
</tr>
<tr>
<td>AC Interaction</td>
<td>2</td>
<td>.053</td>
<td>.027</td>
<td>.965</td>
</tr>
<tr>
<td>BC Interaction</td>
<td>1</td>
<td>.007</td>
<td>.007</td>
<td>.261</td>
</tr>
<tr>
<td>ABC Interaction</td>
<td>2</td>
<td>.003</td>
<td>.001</td>
<td>.054</td>
</tr>
<tr>
<td>SC Interaction</td>
<td>18</td>
<td>.498</td>
<td>.028</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>47</td>
<td>4.148</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For the purpose of presentation and clarity of understanding, each hypothesis will be restated in this section, and then statistically discussed.
Research Hypothesis I

It was hypothesized that there is no significant difference in overall student rating of black and white instructors on the instructor rating form.

The hypothesis presented combines all students ratings for the black instructor and compares them with the same students ratings of the white instructor. This hypothesis allows us to draw conclusions based upon instructor ratings regardless of sex or class level. The specific data relating to Hypothesis I is presented in Table 8.

Table 8. Analysis of variance table relating to Hypothesis I

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructor (C)</td>
<td>1</td>
<td>1.500</td>
<td>1.500</td>
<td>54.279</td>
</tr>
<tr>
<td>SC Interaction</td>
<td>19</td>
<td>.498</td>
<td>.028</td>
<td></td>
</tr>
</tbody>
</table>

There is a significant difference between the two ratings of the instructors. An F of 54.279 is beyond the F required of 4.41 at the .05 level. As there was a significant difference between the two instructors, a review of the mean scores for each instructor was necessary to determine the direction of the difference.

The overall mean score (N=12) for the black instructor was 2.986 as compared to a 2.631 for the white instructor. As
a result the black instructor was rated significantly higher than the white instructor.

Research Hypothesis II

It was hypothesized that there is no significant difference between graduate division, upper division and lower division class ratings of black and white instructors on the instructor rating form.

Hypothesis II was written to determine if students within certain university classifications (graduate, upper and lower divisions) rate black and white instructor differently. Table 9 indicates the results of testing Hypothesis II.

Table 9. Analysis of variance table relating to Hypothesis II

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classes (A)</td>
<td>2</td>
<td>.155</td>
<td>.078</td>
<td>.750</td>
</tr>
<tr>
<td>Subjects (S)</td>
<td>18</td>
<td>1.863</td>
<td>.104</td>
<td></td>
</tr>
</tbody>
</table>

An F value of .750 is nonsignificant at the .05 level. An F with two and 18 degrees of freedom required a 3.55 for significance. Therefore, the null hypothesis was not rejected. The mean scores for the graduate division, upper division and lower division were 2.886, 2.752 and 2.786 respectively with N of 4 in each division.
Research Hypothesis III

It was hypothesized that there is no significant difference between male and female ratings of black and white instructors on the instructor rating form.

Hypothesis III was written to determine if there exists a difference in the ratings of black and white instructors based upon the student's sex. The results of the statistical test for Hypothesis III is given in Table 10.

Table 10. Analysis of variance table relating to Hypothesis III

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sums of Squares</th>
<th>Mean Square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex (B)</td>
<td>1</td>
<td>.032</td>
<td>.032</td>
<td>.311</td>
</tr>
<tr>
<td>Subjects(S)</td>
<td>18</td>
<td>1.863</td>
<td>.104</td>
<td></td>
</tr>
</tbody>
</table>

As indicated in the analysis table an F of .311 is nonsignificant at .05 level. An F of 4.41 is required with one and 18 degrees of freedom. The mean score (N=12) for females was 2.834 as compared to 2.782 for the males.

Research Hypothesis IV

It was hypothesized that there is no first-order interaction between instructor X class, instructor X sex and class X sex on the instructor rating form.
The first order interaction hypotheses were written to determine if any possible combination of effects influenced black and white instructor ratings. The test of significance for each first-order interaction is indicated in Table 11.

Table 11. Analysis of variance table relating to Hypothesis IV

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB Interaction</td>
<td>2</td>
<td>.036</td>
<td>.018</td>
<td>.173</td>
</tr>
<tr>
<td>Subjects (S)</td>
<td>18</td>
<td>1.863</td>
<td>.104</td>
<td></td>
</tr>
<tr>
<td>AC Interaction</td>
<td>2</td>
<td>.053</td>
<td>.027</td>
<td>.965</td>
</tr>
<tr>
<td>BC Interaction</td>
<td>1</td>
<td>.007</td>
<td>.007</td>
<td>.261</td>
</tr>
<tr>
<td>SC Interaction</td>
<td>18</td>
<td>.498</td>
<td>.028</td>
<td></td>
</tr>
</tbody>
</table>

As the results show, none of the first-order interactions proved to be significant. As a result, any two combinations of main effects did not contribute to a difference in black and white instructor ratings.

The three F values for the first-order interactions were .173 for the interaction of class X sex (AB), .965 for class X instructor (AC), and .261 for sex X instructor (BC). The significant level was 3.55 for instructor X class and class X sex with two and 18 degrees of freedom, and 4.41 for instructor X sex, with one and 18 degrees of freedom.
Research Hypothesis V

It was hypothesized that there is no second-order interaction between instructor X class X sex on the instructor rating form.

As with Research Hypothesis IV, the interaction of variables is of concern in Hypothesis V. As there were three major effects, a second-order interaction needed to be tested. Table 12 indicates the results of the second-order interaction test.

Table 12. Analysis of variance table relating to Hypothesis V

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC Interaction</td>
<td>2</td>
<td>.003</td>
<td>.001</td>
<td>.054</td>
</tr>
<tr>
<td>SC Interaction</td>
<td>18</td>
<td>.498</td>
<td>.028</td>
<td></td>
</tr>
</tbody>
</table>

An F value of .054 for the second-order interaction effect reveals that the interaction of instructor X sex X class (ABC) is nonsignificant. An F of 3.55 with two and 18 degrees of freedom is required for significance.
Chapter Summary

From the analysis of the data gathered and the presentation of that data in this chapter, the only major effect found to be significant was the instructor variable. Reviewing the direction of the significance indicated that the black instructor was rated higher than the white instructor. Sex and class level were found to be non-significant and therefore, had no effect upon students' evaluations of black and white instructors. In addition, none of the interaction dimensions proved to be significant.
CHAPTER V. SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

Chapters I through IV of this study has delineated the details of the research project undertaken. This chapter serves as a summary of those details, presents conclusions and list recommendations.

Summary

To summarize this study, it is necessary to describe what has been stated in the preceding chapters. Therefore, a brief description of the study will serve as the summary.

The foundation for which this study is built is taken from two broad areas that are occurring simultaneously in higher education. The growing use of student evaluations of faculty and the increased emphasis on the employment of minorities as faculty members in predominantly white colleges and universities, have given emphasis to this study. Literature cited in Chapter I supports this claim. These two factors led to the statement of the problem of the study.

The problem of the study was to ascertain if skin color made a difference in instructor ratings by predominantly white classes in a predominantly white university. The study also had three distinct purposes.
The three were:

1. To provide the black educator in the predominantly white university with information that he can use to review his student's evaluation of his performance.

2. To provide administrators in higher education with data that will be helpful in reviewing student's evaluations of black instructors.

3. To provide a means of detecting racial discrimination and to initiate programs to eliminate discrimination.

In addition to the problems and purposes, three major and two supportive hypotheses, assumptions, limitations and a procedural outline for the study were written with regard to a number of factors that directly effected this study. The actual procedure used in the study followed a precise plan that was developed around a factorial analysis of variance with a repeated measure. The design was adopted in order that three main effects could be considered.

The next step was to cite research applicable to this study. The chapter entitled Review of the Literature was divided into Four topic areas. The first topic was student evaluations in higher education. The worth of student evaluations along with studies that have at-
tempted to answer many of the questions concerning student evaluations were discussed.

The second area of the review centered upon sex, class level and black-white differences in student ratings of faculty. In the first two areas (sex and class level) a number of studies were discussed which differed in their findings. Some authors found difference in student ratings of faculty based upon sex and class level while others found little or no differences. As there wasn't any data available on students' rating of black and white faculty, a number of studies relating skin color to student evaluations were reviewed.

A theoretical base for student ratings, along with the reliability, validity and use of student ratings was the third area of the review. A theoretical base for present evaluations was established as well as criticized. An alternative theoretical base was also discussed. Most studies have found high reliability of student ratings. However, the validity of the ratings were somewhat questionable. The use of student ratings range from self-improvement to administrative decision making regarding tenure and promotion.

The last area of the review focused upon federal legislation and affirmative action policies that have helped to increase the number of minorities hired as
faculty members at predominantly white colleges and universities. This fourth area established a base for understanding why many minorities are being attracted to the predominantly white college campus. A summary of the review of literature help to tie the chapter together.

The "Research Design" was a detailed analysis of the steps followed in order to complete the study. The instructor rating form used to obtain data, was developed from a number of existing forms. The video tape production, a critical aspect of the research design, was designed in order to control all extraneous variables. Scripts were written that related to all classes, regardless of subject area being studied, sex or age. Two instructors were selected who met the criteria developed for instructor selection.

The tapes were shown twice to a team of three experts. After the first viewing, the tapes were evaluated and recommendations were made by the experts to improve the tapes. After the second showing of the tapes to the experts, their verbal responses were recorded and written responses tabulated. The written responses were used in a 3 x 4 randomized block design and statistically tested. The results of the statistical test indicated that all three experts were not significantly different in their
ratings of the tapes. After the second ratings, the tapes were ready for use in the experiment.

A stratified random sample was drawn from a population of all Spring Quarter 1976 classes at Iowa State University. Of the twelve classes selected, six were from the physical sciences and six from the social sciences. The classes were further stratified on the basis of lower division, upper-division and graduate division class levels. As the sex dimension was of importance in the study, all classes selected had to have a 20-80 ratio of males to females or females to males.

A pilot study helped to systemize the presentation. The instructors of the randomly selected classes were contacted and asked to participate in the experiment.

Eighteen classes were unable to participate or could not meet the 20-80 sex requirement. Once an instructor agreed to participate, the treatment was administered to his or her particular class. The data were collected, using machined scored answer sheets. Individual classroom mean scores were tabulated and then placed into an analysis-of-variance framework for statistical testing.

The findings of the study were based upon the test of the hypotheses. Of the five hypotheses postulated, only one was significant beyond the .05 level. The overall rating of the black and white instructor, treatment C,
was found to be statistically different. Although the results of this statistical difference might be assumed to be in favor of the white instructor, this didn't prove to be the case. The mean scores for both the black and white instructors indicated that the black instructor was rated higher than the white instructor. Sex, treatment A, and class level, treatment B, of the students had no effect upon the instructor's ratings. The first and second order interactions also proved to be non-significant at the .05 level.

This summary serves as a brief description of this research project. With the summary completed, meaningful conclusions can be drawn.

Conclusions

The conclusions of the study will be discussed in terms of the hypotheses tested. The implications of each test and the relationship of the results to the problem of the study will be concluded.

Research Hypothesis I

It was hypothesized that there is no significant difference in overall student rating of black and white instructors on the instructor rating form.
As indicated from the findings, the black instructor was rated higher than the white instructor. The results of this statistical test may appear to be somewhat different than the view that is commonly held. There may be several reasons why the black instructor received higher ratings. White college students may have been more receptive of black faculty than would be expected. Since the students lived in an academic community, they may have had a tendency to be more liberal than other Americans. In fact, college students in an effort to make the black faculty member feel welcome, may have overreacted in their ratings.

Another explanation that might account for the difference in the instructor ratings may be inherent in the research design. As the researcher who administered the treatments was black, the students may have felt an urge to rate the black instructor higher than the white instructor. This may have been manifested in the concept that they felt they were doing the black researcher a favor.

There may be several reasons why this difference did occur, but the two suggested seems to be viable possibilities. Additional research may uncover reasons behind the findings presented as a result of Hypothesis I.
Research Hypothesis II

It was hypothesized that there is no significant difference between graduate division, upper division and lower division class ratings of black and white instructors on the instructor rating form.

The results of Hypothesis II indicated that there was no difference in instructor ratings based upon the students class level. As a result, it was concluded that the students in the different class levels rate the same in regards to their ratings of black and white instructors. Prior studies have been somewhat contradictory as to the effect class level has had upon student ratings of faculty. The findings of this study does little to resolve that contradiction. However, with regard to black and white faculty, this study does indicate that students within graduate, upper or lower class levels do not differ in their ratings.

Research Hypothesis III

It was hypothesized that there is no significant difference between male and female ratings of black and white instructors on the instructor rating forms.

The results of Hypothesis II revealed no significant difference between male and female ratings of instructors. Therefore, it was concluded that male and female students
rate the same in regards to their ratings of black and white faculty. Since an overall mean ratings for each instructor was used in the analysis, it is most difficult to compare the results of this study with other studies that used item differences on particular instructor rating scales as a base for the analysis. As most of the studies reviewed found only slight differences between male and female student ratings with regard to only specific instructor items, it might be concluded that the results of this study is somewhat consistent with other studies. However, in none of the other studies were the students rating black and white faculty.

Research Hypothesis IV

It was hypothesized that there is no first-order interaction between instructor x class, instructor x sex and class x sex on the instructor rating form.

As a result of the statistical test, no first-order interactions were found. It was concluded that no combination of two main effects contributed to a difference in black and white instructor ratings. For example a graduate female class rated black and white faculty the same as an under division male class.
Research Hypothesis V

It was hypothesized that there is no second-order interaction between instructor x class x sex on the instructor rating form.

There was found to be no significant second-order interaction of the three main effects. In conclusion, students sex, class or instructor rated, had no effect upon the black and white instructors ratings. Therefore, all main effects are additive and there are no effects uniquely attributable to the cells of the experiment.

Restatement of the problem

The problem of the study was to ascertain if skin color made a difference in instructor ratings by white classes in a predominantly white university.

It can be concluded from this study that skin color does make a difference in instructor ratings by white classes at a predominantly white university. The direction of difference was contrary to the commonly held view that discrimination if it occurred would be directed toward the black. However, the black instructor was rated higher than the white instructor. University personnel should be aware that racial discrimination can and may be directed toward
the white instructor.

Recommendations

The recommendations presented will be categorized into two areas. The first area concerns recommendations as a result of this research project and secondly, recommendations that might be considered to strengthen or expand this report.

As a result of this research project, it is recommended that student ratings of black and white faculty be examined carefully by all concerned before any final decisions are made. Although it may be a temptation on the part of faculty and administrators in higher education to adjust students' ratings of blacks to compensate for believed racial discrimination, this might not always be the case. In reality, the opposite or reverse may occur. Therefore, the faculty and administrators should make use of any other available data before making any final decisions about the black instructor's ability to teach. To quote Wilbert J. McKeachie (1969b):

My view is that student evaluations can provide useful evidence of teaching effectiveness, but they are best used when interpreted in the context of other evidence (p. 225).

As expressed by McKeachie, the faculty member or
administrator must be aware of other factors that contribute to good teaching. Knowing that student ratings of black faculty can and may be higher than whites of equal ability is valuable data in determining the black instructor's total effectiveness.

As this research project is the first of its kind to test black-white differences in student evaluations of faculty, several recommendations are made to strengthen or expand the data collected in this report. For clarity of presentation, a numerical listing of these recommendations will be used. The recommendations for this study are as follows:

1. A replication of this study to validate the findings and strengthen the conclusions made.
2. A longer exposure to the black and white instructors.
3. A study similar to this one, using a number of geographically located schools throughout the United States.
4. A study similar to this one, using a number of different schools that vary in student enrollment.
5. A study similar to this one, using a number of different types of colleges (e.g., liberal
arts, state supported multi-colleges and private institutions of higher education).

The recommendations made as a result of this study will help to strengthen, expand and validate what has been reported. The recommendations presented are viable and are worth consideration in future research.
BIBLIOGRAPHY


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)


Costin, F. A graduate course in the teaching of psychology: Description and evaluation. *Journal of Teacher Education*, 1968, 19, 425-432.


Hutchison, J. E. Improving evaluation programs through better understanding of the college student and his professor. Chicago, Ill.: American Educational Research Association, 1974 (ERIC ED 091 420).


Lazovik, G. Evaluation of college teachers following feedback from students. Improving College and University Teaching, 1975, 23(1), 34-35.


Shapiro, S. L. & Stein, B. A. *Student evaluations related to frequency of testing*. 1972. (ERIC ED 093 938).


ACKNOWLEDGMENTS

William Blake once wrote "no bird soars too high if he soars with his own wings". No better phrase could be used to describe the researcher. Without the love of my family this dissertation may not have been possible. I am grateful to my wife, Florene, for the dedication and love she has shown me during my years as a graduate student. To my daughters Kim and Angela, I acknowledge the many hours they have spent waiting for daddy to finish his studies.

To Bob Gelina, I owe a great deal of thanks. Not only for his leadership as my major professor but for his enduring friendship. I feel fortunate to have studied under and learned from such a talented individual.

Thanks to Herman A. Ellis and Herbert F. Wedig, the instructors and friends, who devoted much of their valuable time to ensure the quality of the tapes used in this study. In addition, I would like to thank Wan Lee Cheng and Mike Shea for their technical assistance during the tape production.

I owe a great deal of appreciation to my doctoral committee for their help and guidance and especially Dick Warren for his assistance with the statistical analysis of the study. William Wolansky, George Kizer, Arthur
Gowan along with Bob Gelina and Dick Warren, thank you for seeing me through.

Finally, I dedicate this research project to my parents, Sterling and Ruth Yarbrough. Through the years their confidence, love and affection has been abiding.

EGY
APPENDIX A: LETTER TO DR. W. J. McKEACHIE
January 21, 1976

Dr. Wilbert J. McKeachie
Department of Psychology
The University of Michigan
Ann Arbor, Michigan 48104

Dear Dr. McKeachie:

In your text, Teaching Tips: A Guide Book for Beginning Teachers, 1969, you indicated several outstanding instructor rating scales in use today. As you have done extensive work in the field, I'm asking if you might send me names and addresses of persons who have information and normative data on some of the more outstanding instructor rating scales.

I am presently engaged in my dissertation research to determine if black and white instructors are rated equally by predominantly white classes at a predominantly white university. Your help in this matter would be appreciated.

Sincerely,

[Signature]
Earl G. Yarbrough
Graduate Assistant
Industrial Education Department
Iowa State University
APPENDIX B: LETTER FROM DR. W. J. McKEACHIE
3 February, 1976

Mr. Earl G. Yarbrough
Industrial Education Department
Iowa State University
Ames, Iowa
50010

Dear Mr. Yarbrough:

I'm pleased to know that you are doing research on student ratings and hope that you'll send me a copy of your dissertation abstract when you are finished.

With respect to normative data on faculty rating scales, I would suggest that you write to professor John Feldhusen, Educational Psychology, Purdue University, West Lafayette, Indiana, 47906 for data on the Purdue scale, to Dr. James Derry, Department of Psychology, Purdue University, West Lafayette, Indiana, 47906 for normative data on a great many items included in the "Cafeteria" system which Purdue has developed, to Dr. Kenneth Doyle, Measurement Service Center, University of Minnesota, Minneapolis, Minnesota, 55414 for the Minnesota data, to Dr. Grace French Lazovik, Department of Psychology, University of Pittsburgh, Pittsburgh, Pennsylvania, 15260 for the University of Washington and the University of Pittsburgh scales, and to Dr. Lawrence Aleamoni, Director of IRAD, Professor of Educational Psychology, University of Arizona, 507 Math Building, Tucson, Arizona 85721 for the Illinois data.

Good luck with your dissertation!

Sincerely yours,

Wilbert J. McKeachie
Professor of Psychology
APPENDIX C: IOWA STATE UNIVERSITY INSTRUCTOR RATING FORM
INSTRUCTOR RATINGS

Please rate your instructor on the points listed below in order to provide feedback which will help improve his or her instructional techniques.

INSTRUCTIONS:  
A) Do NOT enter your name.  
B) Name instructor in Block "B."  
C) Name course and section in Blocks "E" and "F."  
D) Do NOT code identification unless so instructed.  
E) Start coding in block for Item 1.

ITEMS:

1) I have taken this course:
   1/A To meet a college requirement.  
   2/B Because it is required in my program.  
   3/C Because it is in my major, but an elective.  
   4/D As an elective course not in my major.

   ("SPECIAL" STUDENTS LEAVE ITEM 2 BLANK)

3) My sex is: 1/A Male or 2/B Female.

4) I am taking this course: 1/A For a regular (A-F) grade. 2/B Pass/NP  
   3/C Satisfactory/Fail.

Please use the following five-point scale to rate your instructor. The rating indicates how this instructor compares with all other instructors you have had at ISU.

1/A Far below average (among the lowest 10%).  
2/B Below average (among the next 20%).  
3/C Average (among the middle 40%).  
4/D Above average (among the next 20%).  
5/E Far above average (among the top 10%).

5) ORGANIZATION/EFFICIENCY  
   The instructor was well organized and used class time efficiently.

6) ATTITUDE  
   The instructor was interested in and enthusiastic about teaching this class.

7) STUDENT INTEREST  
   The instructor stimulates student interest.

8) INTERACTION  
   The instructor was respectful, tolerant, and fair with the students.

9) EXPLANATION  
   The instructor explained the course material clearly.

10) EVALUATION  
    The instructor provided clear and appropriate procedures for demonstrating competency in the course.

NOTE: The instructor may ask you to respond to additional questions starting with Item 11. Please write comments or suggestions on the back of the answer sheet.
APPENDIX D: INSTRUCTOR RATING FORM
INSTRUCTOR RATING FORM

Using the answer sheet given you, please mark the instruction items (1-4) and then rate the lecturer (instructor) on the items (5-16) below. The content of the lectures will be rated by another group of students. This data in part will be used to evaluate the effect and suitability of the lecture tapes for the college classroom. Therefore, you are asked to rate each instructor as if he were teaching a class for which you are enrolled. DO NOT INCLUDE YOUR NAME.

INSTRUCTION ITEMS: Mark the correct letters which apply to you on your answer sheet.

ITEM
1. Lecture number: (1)Number one (2)Number two
2. Your class rank: (1)Freshman (2)Sophomore (3)Junior (4)Senior (5)Graduate
3. Sex: (1)Male (2)Female
4. Your overall grade point average: (1)Less than 1.9 (2)2.0 - 2.5 (3)2.5 - 2.9 (4)3.0 - 3.4 (5)3.5 - 4.0

LECTURE RATING ITEMS: Please use the following five-point scale to rate the lecturer (instructor) in comparison to other college lecturers (instructors) you have had or have now.

Mark (1) for FAR BELOW AVERAGE (among the lowest 10%)
Mark (2) for BELOW AVERAGE (among the next 20%)
Mark (3) for AVERAGE (among the middle 40%)
Mark (4) for ABOVE AVERAGE (among the next 20%)
Mark (5) for FAR ABOVE AVERAGE (among the top 10%)

ITEM
5. ORGANIZATION/EFFICIENCY The lecturer was well organized and used lecture time efficiently.
6. ATTITUDE The lecturer was interested in and enthusiastic about the lecture.
7. STUDENT INTEREST The lecturer stimulated student interest.
8. INTERACTION The lecturer was respectful, tolerant and fair in his presentation.
9. EXPLANATION The lecturer explained the lecture material clearly.
10. LECTURE The lecturer spoke adequately enough and stated words clearly.
11. VISUALS The lecturer presented clear and legible visual material and adequately explained them.
12. PRESENTATION The lecturer had an interesting style of presentation.
13. KNOWLEDGE The lecturer was well informed about the material presented.
14. LECTURER The lecturer is a dynamic and energetic person.
15. CLASS ACCEPTANCE The lecturer was well accepted by the class.
16. INDIVIDUAL ACCEPTANCE The lecturer is a person I would like to have taken a class from.
APPENDIX E: LECTURE SCRIPTS - PARTS I AND II
This is Part I of a two-part series entitled YOUTH AND DRUGS IN THE SCHOOLS. Most of the material presented in this and the lecture that follows is taken from a government publication entitled DRUGS IN THE SCHOOLS presented in 1972. My name is _________________.

The first section of YOUTH AND DRUGS IN THE SCHOOLS will focus upon:

(Chart #1)

1. The scope of the problem
2. Drug abuse surveys
3. National drug arrests
4. The American family tragedy

The drug problem in America has reached critical proportions. The youth of today is subject to a greater abundance of drugs than ever before.

In today's society, the chances are substantial that when a parent sends his child to school, he or she is sending that child into a drug-filled environment - an environment where drugs are bought and sold - where there is considerable pressure for students to use drugs.

The availability of drugs on a regular basis in the
school cafeterias, hallways, washrooms, playgrounds and parking lots is astounding. An example:

In a recent study on drugs in the schools, a 17 year old girl who attended a suburban school was asked by the Drug Abuse Committee to purchase narcotics. In just 2 days she had spent $100 on heroin, barbiturates, amphetamines, LSD and marijuana.

The youngsters who became involved in drugs come from all segments of society.

Not only are the youngsters in the schools victims of the drug epidemic but some of them have become the facilitators of the problems. Many of the drug pushers are students that other students look up to and admire. An example:

In Kansas City, the state police arrested a high school youngster for selling drugs. Not only was he popular with the students, but he was also captain of the football team, dated the prettiest girls on campus and drove a new Mustang to school.

In a recent survey done in San Mateo County, California in a school with an enrollment of 1,900 students, 129 or 7% had been selling drugs. Half of the students selling were 15 or 16 years old.

There has been a number of drug abuse surveys that will indicate the problem this nation is facing.
The National Commission on Marijuana and Drug Abuse recently found that 6% of our high school pupils had used heroin (1 1/2 million youths) - 8% or 2 million had used hallucinogenic drugs such as LSD, mescaline and peyote. Five percent tried cocaine, 8% "speed".

The National Institute of Mental Health also surveyed the problem on a national basis. The survey covered 25 schools located in different areas of the country - 9 east coast, 3 in the southeast - 5 in the midwest and the remaining 8 on the west coast. The selection of schools represented ranged from highly affluent suburban schools to economically deprived schools located in the inner city.

While the study reveals that the problem will vary in intensity from school to school, the most interesting note was that every school in the survey had a substantial problem.

A number of state and local surveys further support the findings of the national surveys. Although there are a number of reports on file, the one that stands out and is typical of others will be discussed. The survey was done by a group of sociologists in a junior high school located in a small town in New Jersey. The sample included 400 students from different phases of the societal
structure. The researchers indicated that the 400 students in the study were representative of many of the student bodies found in numerous other small communities across the United States.

(Chart #3)

The results of the study indicated that 15% of these junior high school youngsters had used drugs and 4% of this group had admitted frequent use. One of the frightening findings was that 2 1/2% of the eighth graders in this town of 50,000 people were into heroin. In addition, 31% of these students knew where they could find drugs if they wanted them.

To further demonstrate the drug abuse problem, a look at the national drug arrest statistics is helpful. The Federal Bureau of Investigation reported in 1971 that the national narcotic arrest of youngsters under the age of 19 had skyrocketed 765% over the previous 5 years. In fact, every region of the country has seen this tremendous upsurge in drug prosecutions.

(Chart #4)

The chart here indicates the five states that experienced the highest rate of prosecutions from 1966-1971:
1. West Virginia 1,325% increase
2. Mississippi 1,075% increase
3. Alabama 709% increase
4. Tennessee 679% increase
5. North Carolina 489% increase

These figures only indicate the amount of increase in prosecutions and not the states with the most arrests for drug abuse.

The select Committee on Crime described the drug abuse problem in this nation as the "new tragedy of the American family. It is proliferating and spreading, according to experts, like a contagious disease".

Heroin kills more young people in New York City than any other single cause, including heart disease, cancer, homicides and suicides.

The problem is not unique to New York City. In Los Angeles, California, for the year 1971, two people died each day from an overdose of heroin, while two others commit suicide each day because of an intentional overdose of barbiturates. In Chicago, overdose deaths has increased 50% from 1969 to 1971.

In addition to the deaths caused by an overdose of narcotics, there has been an increase in the number of hospital treatments for overdose victims. At Jackson Memorial Hospital in Miami, Florida, there was reported
as many as five drug overdose cases a day.

The youthful drug problem doesn't only effect the youth themselves, but their families. In testimony before a White House Committee, one mother told of her son's encounter with drugs. Although she knew of her son's problem, as did the proper authorities, he wasn't given proper medical treatment. As a result, the youngster brought greater tragedy to his family when he strangled to death his 5 year old sister.

The drug problem has not only affected the youth of America, but the family from which these youths come. The problems are not limited to "those other people" but to all Americans, whether directly or indirectly.

The scope of the youthful drug problem is great. The youngster in most cases must live with the constant reality that drugs are a part of his youthful environment. It is not a select few or specific group of people who become involved, but the problem often finds its way into a number of unsuspecting communities.

From the surveys cited, it is clear that a youthful drug using society exists. It is also clear that drugs can be located by the youths with little or no effort. Not only the less dangerous drugs, but the "hard core" drugs are finding their way into our schools.

As further evidence of the problem, the national drug
arrest data indicates that there is an increase use of the illegal drugs.

The final area covered dealt with the effect of the youthful drug problem in the lives of the American family. In a number of cases, drugs have been linked to the tragedies of a number of American families. Not only direct death, as a result of an overdose, but death as a result of an individual's drug addiction.
YOUTH AND DRUGS IN THE SCHOOLS

Part II

This is Part II of a two-part series entitled YOUTH AND DRUGS IN THE SCHOOLS. As stated in Part I, most of the data presented here is taken from a government publication entitled DRUGS IN THE SCHOOLS, presented in 1972. My name is ____________________________.

The second part of this presentation will focus upon:

1. The drugs used by the high school students
2. Programs for youth drug users and
3. The school's response

To fully understand the nature of the drug problem, an understanding of drugs used by teenagers is essential. Therefore, the drugs that will be discussed are:

(Chart #2)

1. Heroin
2. Cocaine
3. Barbiturates
4. Amphetamines
5. LSD
6. Mescaline/peyote and
7. Marijuana

Heroin is one of the most dangerous drugs used by school
students. It is perhaps the most addictive drug known. It is white, odorless, and in powered form, is bitter to the taste.

The drug is usually taken by injection or by inhaling. Once taken, it produces a sleepy dreamlike trance.

The heroin user on subsequent occasions must increase the amount of dosage taken to obtain the same high. With these increased injections, the user becomes addicted.

Most of the heroin that reaches the United States is grown in the poppy fields of Turkey, the Middle East and Southeast Asia.

Cocaine is usually referred to as "coke" and is most often found as a white crystalline powder. "Coke" is one of the most expensive of the illegal drugs. The drug is usually taken by inhaling or by injection. It causes general excitement, a feeling of increased physical strength, and a reduction of fatigue. The pulse quickens, blood pressure increases and pupils dilate. Unlike heroin, cocaine doesn't cause strong physical dependence in the abuser, but does create a sense of psychological dependence.

Most of the cocaine found in the states comes from South America. The coca plant grows mainly in Peru and Bolivia.

The barbiturates that are taken by our school-aged youngsters are produced by various pharmaceutical companies.
These barbiturates are usually called by a number of different street names such as "barbs", "downers", "red birds", and "yellow jackets".

In normal doses, these drugs usually depress the actions of the user, lower his blood pressure, slow down heart beat and breathing. In large amounts, it may cause confusion and slurred speech. Like heroin, barbiturates are physically addicting. Taken off of barbiturates abruptly, the user may suffer cramps, nausea, hallucinations and even death!

Amphetamines, another of the drugs often found on the school campus, is produced largely by pharmaceutical companies. They are usually in capsules and tablet forms and can be found in a number of different shapes, sizes and colors. They are often called "footballs", "greenies", "hearts" and "bennies", by the users.

The drug is usually taken orally, but is sometimes dissolved and injected. When misused, the drug causes excitement, restlessness, talkativeness, insomnia and tremor of the hands.

LSD is usually referred to as "acid" and is an extremely powerful hallucinogen drug. It is colorless, odorless, tasteless and is usually taken orally.

When taken, LSD often causes dilated pupils, lower temperature, shivering, chills and perspiration.
Mescaline and peyote are also hallucinogen drugs but are less powerful than LSD. Peyote is usually found in dry, leather-like buttons cut from cactus plants which bloom in Mexico and the Southwestern United States. The buttons are chopped, ground and placed in capsules or rolled in small balls.

Mescaline is derived from peyote cactus buds and is a solid brown green color in its natural form. When distilled, it becomes a clear liquid.

Marijuana is derived from the flowering tops, leaves and small stems of the cannabis plant. These parts are collected, chopped like tobacco and smoked. The cannabis plant grows best in hot and dry climates.

The physiological and psychological effects of marijuana vary from person to person and/or from time to time. Death from cannabis overdose appears to be extremely rare.

There are a number of rehabilitation centers located throughout the states to help the youthful drug user. Two of these programs will be discussed briefly.

(Chart #3)

1. The seed and

2. Gateway House Foundation, Inc.
The Seed program is located in Fort Lauderdale, Florida. The program is an intensive three-week group therapy program that is followed by three months of out-patient involvement.

The youth are referred to the Seed program from a number of sources - relatives, friends, police officers, etc. After being accepted, the youngster participates in group therapy on a daily basis for three weeks. The sessions are conducted by young people who have completed the program. The sessions last from 10 o'clock in the morning until 10 o'clock at night. These young drug abusers are not allowed to return to their homes at night, but are scheduled to stay with the family of youngsters who have completed the same program.

While with the Seed program, the youngster finds that his problem is not unique! Others also have suffered and are there to help the participant kick the habit. One of the successes of the program seems to be the youth counselors.

The Gateway House Foundation maintains six separate treatment facilities in and around Chicago for approximately 200 residents. Each of these facilities is a part of an integrated "therapeutic community" approach to treatment of drug addiction.

The therapeutic approach places the drug addicts in
small, highly controlled, structured, residential settings. The goal is to detoxify the addict so that he can function in a drug-free environment.

The first step in the program is the induction phase. To be admitted to this phase, an addict must have demonstrated a desire to kick the drug habit. This first phase is 60 days long. After the first 60 days, the addict is assigned to one of the programs' three major facilities where an intensive rehabilitation effort begins. Each of these three centers have a twenty-four-hour day, live-in, work-in program. This phase of the program lasts ten months.

The final stage of the program is called re-entry. Re-entry into society is carefully controlled in a facility which performs a function similar to the halfway houses. The participants continue to reside at a Gateway House, but either work outside the facility or attend school.

These two programs are only a few of the many in operation today. Although these and other programs vary in structure, they all seem to have similar concepts.

Many school districts are doing little or nothing in the way of confronting drug abuse. The policy of the school in a lot of cases seem to be to ignore the problem.

Although many schools have become a sanctuary for drug sales, these schools have failed to deal with the
problem effectively. In some school districts, tremendous confusion exists as to the role of the school, administration and teachers in dealing with youthful drug offenders. In addition, most teachers are unable or unwilling to cope with students high on drugs.

Most of the schools that are attempting to eliminate the problem have poorly conceived and often counterproductive policies on drug abuse.

The schools in most cases are not to blame for this situation. Often, funding for drug abuse programs is limited or nonexistent. In many major school districts, the entire drug program has been placed under the direction of one person. The financial support in some schools for drug education is less than 5 cents per child per school year.

From our discussion in Part II, a description of many of the drugs used by our school-aged youngsters was given and explained. Hopefully, this knowledge is helpful in understanding the drug problem facing our nation.

Another aspect of this lecture dealt with programs for young drug users. Although there are many rehabilitation programs, there still isn't enough being done. The two programs mentioned in the lecture are representative of many outstanding programs in existence.
The schools, in most cases, are inadequate in dealing with the drug problem. There are a number of good reasons: the small number of professional staff involved, the amount of money spent per child and the lack of drug education for teachers. In the final analysis, little is being done to eliminate the growing drug problem.
APPENDIX F: LETTER FROM DR. D. J. ZAFFARANO
March 17, 1976

To: Faculty Colleagues

From: D. J. Zaffarano

Mr. Earl G. Yarbrough has shown me his dissertation proposal involving a study of Black-White attitudes at ISU and has asked for a letter of endorsement and support. My understanding is that his committee, consisting of Drs. Wolansky, Gelina, Warren, Kizer, and Gowan have approved the proposal.

Mr. Yarbrough will request the cooperation of several classroom instructors in order that he can administer his questionnaire to students in a random sampling of classes.

Assuming that the research design is adequately considered, I agree that the results of this experiment would be interesting. I do not advocate that this test period should replace a regular class meeting, but if a mutually satisfactory way can be devised to permit your students to assist Mr. Yarbrough in acquiring his data, I am sure he will be very grateful.

DJZ/mhl