

1974

An Experimental Study of the "Doctor Fox Effect"

John E. Ware
Southern Illinois University

Reed G. Williams
Southern Illinois University

Follow this and additional works at: <https://scholarworks.uni.edu/istj>



Part of the [Science and Mathematics Education Commons](#)

Let us know how access to this document benefits you

Copyright © Copyright 1974 by the Iowa Academy of Science

Recommended Citation

Ware, John E. and Williams, Reed G. (1974) "An Experimental Study of the "Doctor Fox Effect"," *Iowa Science Teachers Journal*: Vol. 11: No. 3, Article 5.

Available at: <https://scholarworks.uni.edu/istj/vol11/iss3/5>

This Article is brought to you for free and open access by the IAS Journals & Newsletters at UNI ScholarWorks. It has been accepted for inclusion in Iowa Science Teachers Journal by an authorized editor of UNI ScholarWorks. For more information, please contact scholarworks@uni.edu.

Offensive Materials Statement: Materials located in UNI ScholarWorks come from a broad range of sources and time periods. Some of these materials may contain offensive stereotypes, ideas, visuals, or language.

AN EXPERIMENTAL STUDY OF THE "DOCTOR FOX EFFECT"

John E. Ware, Jr. and Reed G. Williams
Southern Illinois University
Carbondale, Illinois

Colleges and universities are employing student evaluations of faculty with increasing frequency. On some campuses, the forms are used at the instructor's option and serve primarily as a source of feedback for faculty development purposes. Others require student evaluation of instruction and publish the results in academic "consumer's guides." Still others consider student evaluations of instruction in making decisions with regard to faculty salary increases, retention, tenure and promotion. The optional use of student ratings in order to improve instruction often evolves to the mandatory consideration of student rating data in making decisions about faculty retention, promotion and tenure. As this occurs it is important that the validity of these measures also increases. Unfortunately, however, not much is known about what student ratings of faculty really measure and experimental studies which can extend this knowledge base are virtually non-existent.

In establishing the validity of a student-faculty evaluation questionnaire, one must answer two questions. First, is the evaluation instrument sensitive to differences in instruction? Second, does the instrument provide a rating which is a valid index of overall instructional effectiveness? The second question is important in that the student evaluation of instruction is used not only as a direct measure of student opinions regarding the instructional experience but as a presumably adequate substitute for at least one other direct measure of teaching effectiveness; student academic achievement.

Many have argued that students are the most qualified to evaluate teaching effectiveness. Their position is summarized nicely by the following quotation attributed to Aristotle:

You get a better notion of the merits of the dinner from the dinner guests than you do from the cook.

However, examination of the published literature, primarily correlational results, regarding the validity of student ratings of teaching effectiveness yields inconsistent evidence. For example, Costin, Greenough and Menges and Royce note the argument that student ratings of faculty may reflect nothing more than student judgments of the

entertainment value of instruction. Guthrie on the other hand claims that popular teachers are also "substance teachers." The correlational evidence regarding the relationship of student ratings to learning gain is not consistent. Even in those instances where correlational findings are somewhat consistent, the evidence does not permit the luxury of understanding which, among the variables compared, are the "chickens" and which are the "eggs."

Carpenter and Haddan described anecdotal evidence regarding the responses of students to a lecturer who presented vague generalities on a topic unfamiliar to him in a deliberate and confident manner. The speaker, who was introduced as an expert, was applauded and many wanted to invite him to return. When asked to evaluate the speaker, the students were nearly unanimous in his favor even though they confessed later that they could not recall anything of an informative nature that had been said. Based on this experience, Carpenter and Haddan suggested that levels of information-giving and emotional delivery be manipulated in order to determine effects on students.

Coats and Smidchens programmed lecturers to deliver a verbatim script with variations in levels of dynamism (use of gestures, vocal inflection, animation and so on). They reported that listeners who viewed dynamic presentations learned more than did listeners who viewed less dynamic presentations. However, they did not secure ratings of instructional effectiveness, so it was not possible to determine whether the differences in student achievement were reflected in ratings of instruction.

More recently, in a controversial article published in the *Journal of Medical Education*, Naftulin, Ware and Donnelly questioned the validity of student ratings claiming that they may only reflect the seductive style of the instructor. They hypothesized that, given a sufficiently impressive lecture paradigm, students could be "seduced" into feeling satisfied they had learned despite the presentation of irrelevant, conflicting and meaningless content. An actor was programmed to present a topic by way of lecture. He was coached to make considerable use of double talk, neologisms, nonsequiturs and contradictory statements. All of this was placed in the context of seductive gestures, parenthetical humor and meaningless references to unrelated topics. In short, the speaker who was introduced as "Doctor Fox" gave a very enjoyable lecture to a live audience in which he offered little or nothing of substance.

Subjects in the "Dr. Fox" study evaluated Dr. Fox's lecture favorably on an eight-item satisfaction questionnaire. These findings were taken as evidence that students and faculty may evaluate lectures favorably even in the absence of substance in lecture presentations. This phenomenon was characterized as an "illusion" of having learned. However, the listeners were not asked to rate learning gain and no measure of achievement was employed. Therefore, it is not possible to know whether or not an illusion actually occurred.

The present study was designed to provide an experimental test of the effects of lecturer seduction and content-coverage on student ratings of instruction and student achievement. Specifically, the authors sought to determine if: 1) content-coverage and seduction interact with each other in affecting student ratings or test performance, 2) seduction independently affects these student outcomes, and 3) content-coverage independently affects these student outcomes.

METHOD

Subjects

The subjects were 280 undergraduate and graduate students who were enrolled in three general studies sections at Southern Illinois University (SIU) during the fall quarter, 1973. Analyses reported were based on 207 students who attended classes on the day the study was conducted and for whom complete data were available. Thirty-three percent of the subjects were males and 67 percent were females. They ranged in age from 17 to 42 years with a median age of approximately 20 years. Twenty-one percent were freshmen, 30 percent were sophomores, 28 percent were juniors, 18 percent were seniors and 3 percent were graduate students. The largest proportion of students (42 percent) was enrolled in liberal arts and sciences. Other academic majors included education, engineering and technology, business and home economics.

Procedure

An experimental design was used in which six lecturer types were studied. These lecturer types were achieved through the use of six verbatim scripts and the production of six videotaped lecture presentations. All lectures covered the same topic, the biochemistry of memory. The lecturer was the Hollywood actor who was "Dr. Fox" in the studies reported by Naftulin, Ware and Donnelly. The lectures were systematically varied in terms of number of substantive teaching points covered and degree of seduction in doing so.

In the high, medium and low content lectures, 26, 14 and 4 sub-

stantive teaching points were covered, respectively. When substantive teaching points were removed, experimental details without results were substituted along with discussion of unrelated examples and short stories, discussion of what "was going to be covered," and circular discussions of unimportant or meaningless thoughts. Three lectures (high, medium and low content) were delivered in a seductive manner while three lectures (one at each content level) were delivered so as to be low in seduction. The operational definition of seduction included a number of lecturer characteristics described in the published literature which are independent of content-coverage. These behaviors appear likely to influence whether or not a lecturer persuades or entices his audience with regard to the subject matter. High seduction was associated with enthusiasm, humor, friendliness, expressiveness, charisma and personality. One or more aspects of this dimension have been described in each of several factor analytic studies of teacher behavior. A complete description of the production of the videotaped lectures and other procedures followed in the current investigation is available elsewhere.

Groups were formed by dividing each of three existing class groups using a table of random numbers. Lecture conditions were assigned to the six groups using a table of random numbers. One videotaped lecture was shown to each group of students. The lecturer was not introduced. Students were asked to evaluate the lecture presentation using an 18-item student-faculty rating questionnaire. Items pertained to a variety of lecturer behaviors and student outcomes such as increased knowledge, inspired confidence, broadened interest and increased appreciation for the subject. Students were also asked to complete a 26-item multiple choice test over the lecture topic.

Analyses

An unweighted means analysis of variance was used to test hypotheses about interactions and main effects for cognitive test scores and student satisfaction ratings. The analysis procedure is an application of the least squares criterion to the general linear model. It is a statistical solution to analysis of variance designs with unequal cell frequencies in which treatment groups were formed from natural groups differing in size but not differing in other characteristics. The assumption of equivalence of lecture groups prior to viewing the lectures was met for all seven subject characteristics for which data were available (sex, age, GPA, academic standing; and *a priori* ratings of knowledge

of the subject, interest in the subject and expected grade in a course on the subject).

RESULTS

Mean rating and cognitive test scores for the six treatment groups are shown in Table 1. The unweighted means analysis of variance yielded no significant interaction between seduction and content main effects for the cognitive test scores ($p > .05$, $F=1.6$, 2 and 201 df). However, both content ($p < .01$, $F=5.7$, 2 and 201 df) and seduction ($p < .01$, $F=21.7$, 1 and 201 df) main effects were significant among the six groups. Differences in test performance are shown in Figure 1. Students who viewed lectures high in seduction performed better on the cognitive test than did students who viewed the same lectures delivered so as to be low in seduction. Likewise, students who viewed lectures high in content performed better on the cognitive test than did students who viewed lectures lower in content.

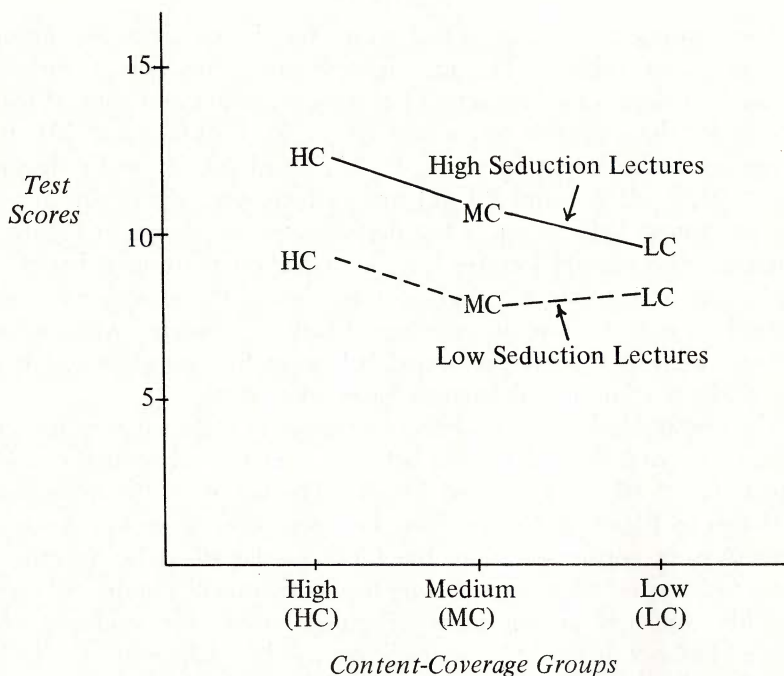
The unweighted means analysis of variance of satisfaction ratings indicated a significant interaction between content and seduction main effects ($p < .05$, $F=4.3$, 2 and 201 df). The nature of the interaction is shown in Figure 2. Comparisons between pairs of group means of interest were performed using the formulas described by Scheffe in order to interpret differences among the six treatment groups. All three possible contrasts among pairs of group means for students who viewed lectures differing in content-coverage but delivered so as to be high in seduction were not significant (high content-high seduction vs. medium content-high seduction, $F=.00$; high content-high seduction vs. low content-high seduction, $F=.70$; medium content-high seduction

TABLE 1
MEAN SATISFACTION RATING SCORES AND
COGNITIVE TEST SCORES

Score	<i>High Seduction Groups</i>			<i>Low Seduction Groups</i>		
	<i>High Cont</i>	<i>Med Cont</i>	<i>Low Cont</i>	<i>High Cont</i>	<i>Med Cont</i>	<i>Low Cont</i>
Cognitive Test						
Mean	12.5	10.9	9.5	9.5	7.7	8.4
Standard Deviation	5.3	3.4	2.7	3.9	2.6	2.2
Satisfaction Rating						
Mean	58.0	59.1	53.9	48.7	35.5	34.4
Standard Deviation	13.9	15.2	13.7	15.2	11.5	11.5

FIGURE 1

PLOT OF MEAN TEST SCORES

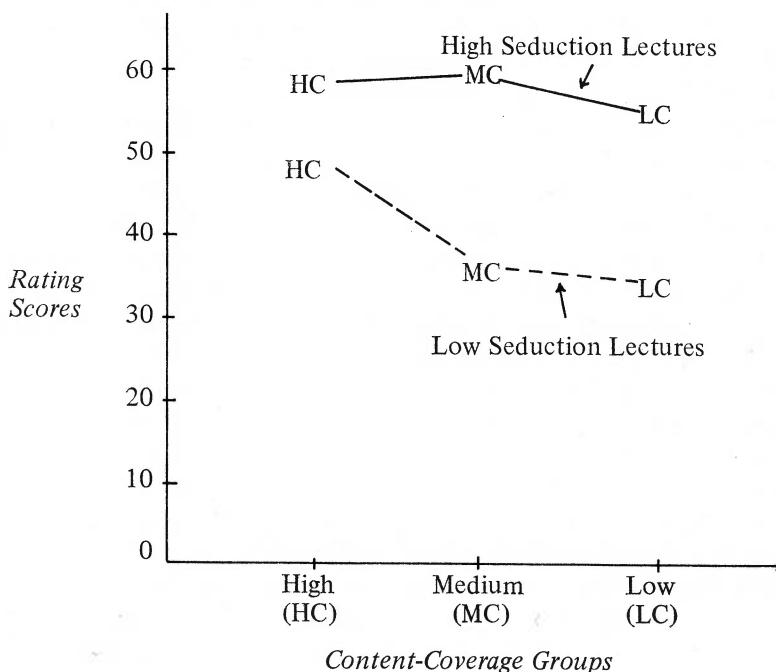


vs. low content-high seduction, $F=.90$; $p > .05$, $df = 5$ and 201 for all F tests). One of the three possible comparisons between pairs of group means for the three content groups who viewed lectures delivered so as to be *low* in seduction were significant. Students who viewed the high content-low seduction lecture gave higher ratings than those who viewed the medium content-low seduction lecture ($p < .05$, $F=12.6$, 5 and 201 df).

Four additional contrasts among pairs of lecture group means were performed in order to determine whether or not differences in student ratings exist between high and low seduction lectures at each of the three content levels. Differences between high and low seduction groups were significant at both the medium content level ($p < .05$, $F=25.6$, 5 and 201 df) and at the low content level ($p < .05$, $F=12.5$, 5 and 201 df). Thus, students tend to rate lectures delivered so as to be high in seduction more favorably than lectures delivered so as to

FIGURE 2

PLOT OF MEAN SATISFACTION RATINGS



be low in seduction. In order to provide an extreme test of the influence of seduction as opposed to content-coverage, one additional contrast of group means was performed. The mean rating score for the low content-high seduction group was compared with the mean rating score for the high content-low seduction group. The mean difference was not significant although the high seduction-low content group tended to give higher ratings.

CONCLUSIONS

Both content-coverage and seduction appear to affect student ratings and student test performance. The influence of these lecturer characteristics on student test performance was more straightforward than their influence on student ratings. Student test performance was higher with increased levels of content-coverage in a lecture presentation. Also, lectures higher in seduction produced higher test scores

than the same lectures delivered so as to be low in seduction. In terms of total variance accounted for, lecturer seduction appears to be the more important influence on test performance of the two characteristics studied in this investigation. The findings with regard to effects of lecturer seduction on student test performance are consistent with those published by Coats and Smidchens.

Content-coverage and seduction in a lecture presentation appear to interact with each other in affecting student ratings. However, the interaction was ordinal with higher mean rating scores for students who viewed high seduction lectures than for students who viewed low seduction lectures. The observation that students give favorable ratings under conditions of high seduction regardless of level of content-coverage is consistent with the observations of Carpenter and Haddan; and Naftulin, Ware and Donnelly.

The results of the present study suggest that the ratings of students in the three content groups exposed to low seduction lectures more accurately reflect differences in content-coverage. Students in groups who viewed lectures higher in content-coverage tended to give higher ratings than those who viewed lectures lower in content-coverage. Thus, it appears that student ratings of a non-seductive "Doctor Fox" may be more valid in terms of the content-coverage criterion.

Under high seduction lecture conditions, student ratings were not sensitive to differences in content-coverage. No differences in ratings were observed for groups differing in content-coverage under high seduction conditions. In the most extreme comparison, the mean ratings of students who viewed a lecture covering 26 substantive teaching points did not differ from the ratings of students who saw a lecture covering only four substantive teaching points. Given that the three content groups under high seduction actually differed in test performance but not in student ratings brings the validity of the latter into question in terms of the content-coverage criterion. It appears that student ratings of teaching faculty may, under all conditions, reflect the influence of seduction, i.e., the "Doctor Fox Effect." When seduction is high, student ratings do not seem to reflect differences in content-coverage.

It would seem that the use of student ratings to improve teaching and to make decisions regarding faculty may be a double-edged sword. Faculty who master the "Doctor Fox Effect" are more likely to receive favorable student ratings regardless of how well they know their subjects and regardless of how much their students learn. This problem

can be avoided with inclusion of a direct measure of student achievement along with student ratings in a faculty evaluation program.

Certain other questions related to the study of lecture presentations and the validity of student ratings remain to be answered. The extent to which the differences in test performances observed in this study are due to differences in learning and/or test-taking motivation requires further study. It is conceivable that lecturer seduction may act on one or both of these capabilities. A study designed to address this question is currently being conducted by the authors.

Other methods of computing student rating scores should be investigated. Preliminary analyses by the authors in which a weighted linear combination of rating scale items was used to discriminate among different lecturer types suggest that such scoring procedures may improve the usefulness of student ratings.

The phenomenon of educational seduction which has come to be known as the "Doctor Fox Effect" appears to be more complicated than was originally thought. Naftulin, Ware and Donnelly were correct in anticipating that teaching effectiveness may be optimized with high levels of content-coverage *and* a seductive presentation manner. On the other hand, the "Doctor Fox Effect" appears to be much more than an illusion. Whereas teaching style appears to be a major factor in determining student ratings, it is also a powerful influence on student test performance.

INFORMATION ABOUT THE NATIONAL YOUTH SCIENCE CAMP

In 1963 the state of West Virginia founded the National Youth Science Camp and the governor of that state invited each of the other governors to nominate two senior boys as delegates, all expenses paid by the state of West Virginia. This year both delegates and the alternates are being selected on behalf of Governor Ray from the Hawkeye Science Fair. These individuals will spend almost three weeks in late June and early July at the camp high in the West Virginia mountains studying and having fun. The delegates this year are Dean Loven, Senior High School, Newton, and Randy Stalzer, Garrigan High School, Algona. The first alternate is Marc Johnson, Community High School, St. Ansgar, and the second alternate is Tim Hanusa of Abraham Lincoln High School, Council Bluffs.