Proceedings of the Iowa Academy of Science

Volume 57 | Annual Issue Article 55

1950

# A Forced-Choice Rating Scale for College Instructors

A. R. Rustebakke
*Grinnell College*

George D. Lovell
*Grinnell College*

Charles F. Haner
*Grinnell College*

## Recommended Citation

Rustebakke, A. R.; Lovell, George D.; and Haner, Charles F. (1950) "A Forced-Choice Rating Scale for College Instructors," *Proceedings of the Iowa Academy of Science, 57(1),* 399-406.
Available at: https://scholarworks.uni.edu/pias/vol57/iss1/55

# A Forced-Choice Rating Scale for College Instructors

By A. R. Rustebakke, George D. Lovell, and Charles F. Haner

## Purpose and Introduction

The purpose of this study was to construct a forced-choice rating scale for evaluating students' opinions of college instructors. The scale constructed consists of 20 tetrads, or groups of 4 statements descriptive of instructors. These statements were chosen so that a pair of favorable items, both appearing to be equally favorable, and a pair of unfavorable items, both appearing to be equally unfavorable, make up each tetrad. From each tetrad the rater must choose the item most characteristic and the item least characteristic of the ratee.

The reason for choosing items in this manner can be seen in Sisson's statement of the basic assumptions underlying the forced-choice method, which was used in rating Army officers. He says that "the basic assumptions underlying the method can be stated as follows:

> "1. Any real differences which exist between officers in competence or efficiency can be described in terms of objective, observable items of behavior.
>
> "2. These 'behavior items' differ in the extent to which people in general tend to use them in describing other people, i.e., in general favorableness, and this tendency can be determined statistically. (The index of general favorableness is the preference index.)
>
> "3. These item also differ in the extent to which they characterize officers at one extreme of the true scale of competence as opposed to officers at the other extreme. The index of this difference, the 'discriminitive' value, can also be determined statistically. (This index, an indication of the degree to which an item discriminates between good and poor performance, is the discrimination index.)
>
> "4. Pairs of items can be selected such that they are equal in preference value, but different in discriminative value. A rater forced to say which item is most (or least) characteristic of a ratee is thus unable to select solely on the basis of prejudice for or against him (since the preference values are equal). The rater is compelled to consider both alternatives and — theoretically at least — do a more objective job of reporting." (Sisson, 1948)

According to the fourth assumption, a pair of items can be chosen so that they appear to be equally favorable (or unfavorable), but so that one item discriminates between good and poor performance to a greater degree than the other. When a rater is forced to choose one of the two items as being most (or least) characteristic of a ratee, he cannot, since the two items appear to be equally favorable (or unfavorable), give a high (or low) rating solely on the basis of his prejudice for or against the ratee. In other words, the forced choice method reduces the rater's control over the final rating he will give.

## Procedure

Using these same basic assumptions, the construction of the present scale was begun. The first step in constructing the scale was to secure items pertinent to students' opinions of instructors. To obtain such items, some 200 essays written by seniors of Grinnell College were obtained. Half of these were about "best" professors and half were about "worst" professors. All departments and divisions of the college were represented in this return, and since the percentage returned from each division was about the same, it was felt that the returns would be satisfactory for obtaining the items. These essays were carefully read and re-read, and a rough classification of items was set up. Items were classified as to their meaning, and the frequency with which each item so classified appeared was tabulated. The reliability of this classification was not determined. One hundred and seven of the most frequently mentioned items, 53 of which were descriptive of "best" professors, and 54 of which were descriptive of "worst" professors, were chosen to make up the list which was administered as the second step in the construction of the scale.

This list (which will be known as Form 1) was constructed so as to enable the students to rate 3 professors. They did this by merely checking those items they "felt sure" applied to the professors they were rating. The subjects were instructed to choose 3 specific professors with whom they had had experience: one the best professor they had ever had, another an average professor, and a third the worst professor they had had. Thus an overall rating was obtained, which was in terms of these 3 categories. These categories — best, average, and worst — were not further defined; rather it was left up to the student to define them, for the purpose of the scale was to get at what students consider to be best and worst professors. This also controlled the students' choice so that the professors they rated as best constituted the upper portion and those they rated as worst the lower portion on a continuum of students' opinions of teaching effectiveness. The instructions which accompanied Form 1 were worded so as to create a set to combat the "halo" effect which would perhaps be induced by using the terms best, average, and worst, in describing the professors to be rated.

Copies of Form 1 were sent to the 248 seniors of Grinnell College, and 110 of these were returned giving a total of 330 ratings. Since the percentages of all seniors in each of the divisions of the college returning Form 1 ranged from 41-50%, there is some evidence to show that the returns were fairly representative of the group to whom it was sent.

From the 330 ratings on the returns of Form 1, two indices — the preference index and the discrimination index — were determined statistically for each of the 107 items. The preference index is a measure of the frequency with which an item is chosen as applying to the total group to be rated. The following formula was devised to determine the preference indices of the items: Preference index $= \frac{1f_1 + 2f_2}{330} \times 100$, where 1 is the weight of the degree of application "applies," $f_1$ is the frequency with which the item was checked as applying to all three groups — best, average, and worst professors, 2 is the weight of the degree of application "does not apply," $f_2$ is the frequency with which the item was not checked — indicating it did not apply — to all three groups, and 330 is N or the total number of professors rated. Using this method, the preference indices of the items have a possible range of from 100 to 200. A low index indicates that an item is chosen frequently as applying to best, average, and worst professors, is generally favorable, and students tend to use it in describing college instructors. A high index indicates that the opposite is true. Although not a necessity of the method, it was found in this study that the items with a low preference index were those items which are "the nice thing to say" about a professor.

In the determination of the discrimination indices of the items an item analysis technique using biserial correlation was employed. Three assumptions seemed to indicate that this technique was applicable: (1) that the variable, judged effectiveness of teaching, is continuous and normally distributed, (2) that the variable, degree of application of an item, is continuous and normally distributed, and (3) that the best and worst professors as chosen by students constitute the extremes of the variable, judged effectiveness of teaching. If these assumptions are, as they appear to be, valid, then the use of the item analysis technique is justified.

A table, reported by Flanagan to give satisfactory approximations of biserial coefficients, was used in computing the correlation. (Flanagan, 1939) The table is computed for the 27% scoring highest and lowest on the continuous variable — in this study the variable, judged effectiveness of teaching. The use of a table computed for the highest and lowest 27% of a distribution cannot be justified since there was no overall rating of any one professor in this study, but the assumption seemed reasonable that being chosen as a best professor was analogous to obtaining an overall rating in the highest 27% and that being chosen as a worst professor was analogous to obtaining a rating in the lowest 27%. The correlation coefficients computed using this table were regarded as sufficiently accurate measures of the relationship between having an item checked as applicable and occupying a position on the best end of the continuum so that it could be used as the discrimination index. A plus correlation indicates that there is a relationship between being chosen as a best professor and the particular item being applicable; a minus correlation indicates that there is a relationship between being chosen as a worst professor and the particular item being applicable.

Forty pairs of items were then selected so that their preference indices were as similar as possible, their discrimination indices as widely differing as possible, and so that the items chosen completely covered the content of the items in Form 1. Pairs of items were combined to make up the 20 tetrads of the final scale. The pairs were combined so that a favorable pair of items and a mildly unfavorable pair, or a mildly favorable pair and an unfavorable pair were in each tetrad. For example, one tetrad might contain pairs with preference indices of 110 and 150, while another tetrad might contain pairs with preference indices of 140 and 180. The items were chosen so that items having opposite meanings were not placed in the same tetrad, for this would eliminate the "forced-choice" element from the scale. The four items were then arranged in alphabetical order, and the twenty tetrads so arranged make up the final scale on which the rater is forced to choose from each tetrad the one item he thinks most characteristic and the one item he thinks least characteristic of the instructor being rated.

In scoring the scale, only the two most discriminating items, i.e., the two items having the highest positive and negative correlations with being chosen as a best professor, are scored. Using this method of scoring, the maximum possible score is a plus 40 and the minimum possible is a minus 40, a high plus score indicating a best professor and a high minus score indicating a worst professor.

## Results and Interpretation

The chief result of this study is the scale, a copy of which is appended to this paper. A study of the reliability of the scale is at present underway, but as yet is not completed. The test-retest method is being used. In order to eliminate consistency of responses on the retest due to retention of responses on the first test, the students being used as subjects rate 3 or 4 different professors during each of the testing periods. The elapsed time between test and retest was approximately 2 weeks. A tentative reliability coefficient of .912 having a standard error of .028 has been computed. This coefficient is based on only 36 subjects, and is therefore subject to revision as more ratings become available.

As to the validity of the scale, it should be valid in terms of the sample and method of construction on which it is based. The scale is designed to evaluate students' opinions of college instructors. Both the items themselves and the data from which the preference and discrimination indices were computed were drawn from samples of students' opinions. Therefore, the evaluations of students' opinions made by using the scale should be valid. However, a further check on the validity of the scale is being made along with the study of the scale's reliability. No results of this part of the study are at present available.

Interpretation of the results given by the use of the scale requires that the sample on which the scale is based and certain idiosyncrasies of the forced-choice method of rating be considered. First, the scale is so constructed that the overall rating is in terms of the opinions of a sample of Grinnell College seniors, and is subject to the usual limitations of a limited sample. The scale was not constructed to represent the opinions of administrators, educators, or any other group as to the characteristics of a good instructor, but rather to represent student reactions. Second, the forced-choice method of rating forces the rater to choose two of four items, one as most characteristic and one as least characteristic of the ratee. This introduces the possibility that in one of the tetrads there may be two items which are highly characteristic of the person being rated, while in another tetrad none of the items may be highly characteristic. Thus an analysis of the content of the items checked might not give a truly accurate picture of the professor being rated. However, the overall rating should be valid and will not be affected to a very great extent by the bias of the person doing the rating. Inasmuch as the rater can recognize which items are favorable and which are unfavorable, he can control whether the overall rating will be in general favorable or unfavorable, but even if he is able in all cases to recognize the two items in each tetrad that are favorable, and checks one of the two as being most characteristic, since he does not know which of the two items discriminates between best and worst professors, he cannot control how high a favorable rating he will give. This is the chief advantage of the forced-choice method of rating.

## Grinnell College Faculty Appraisal Scale
## Student Form, Experimental

INSTRUCTIONS: The following scale is made up of 20 groups of items, each group being made up of 4 items which are relevant to teaching ability. In each of the groups of 4 you are to choose the one item which you think is most characteristic of the professor you are rating and the one item which you think least characteristic of him. You must choose two items in each group, one of them least characteristic and one of them most characteristic. Place a check in the first column, headed by the word "Most," opposite the item you wish to choose as most characteristic and a check in the second column, headed by the word "Least," opposite the item you wish to choose as least characteristic. Place your check-marks carefully in the center of the parentheses in each column so that accurate scoring will be possible.

| Most | Least | |
|---|---|---|
| ( ) | ( ) | 1. a. Does not use drop quizzes or oral quizzes to keep students up on daily work. |
| ( ) | ( ) | b. Furnishes no yardstick which student may use to judge progress in course. |
| ( ) | ( ) | c. Gives assignments, dates for tests, etc., well in advance. |
| ( ) | ( ) | d. Has thorough knowledge of subject. |
| ( ) | ( ) | 2. a. Difficult to take notes from lectures. |
| ( ) | ( ) | b. Does not read his lectures. |
| ( ) | ( ) | c. Lectures a rehash of assigned material. |
| ( ) | ( ) | d. Presents good appearance (neatness, dress, etc.) |
| ( ) | ( ) | 3. a. Does not integrate course with other fields. |
| ( ) | ( ) | b. Does not "ride" or ridicule students. |
| ( ) | ( ) | c. Is anxious to get material across to students. |
| ( ) | ( ) | d. Speaking technique poor (voice, gestures, etc.) |
| ( ) | ( ) | 4. a. Does not consider student interest in presenting material. |
| ( ) | ( ) | b. Gives enough tests, assigns enough papers, to determine grade. |
| ( ) | ( ) | c. Prepares sufficiently for classes. |
| ( ) | ( ) | d. Students' other activities not considered in assigning papers, tests, etc. |
| ( ) | ( ) | 5. a. Does not explain his grading system. |
| ( ) | ( ) | b. Has the respect of his students. |
| ( ) | ( ) | c. Is willing to admit when he doesn't know something. |
| ( ) | ( ) | d. Reaches no conclusions, presentation so indefinite as to leave student up in air. |

Most Least

( ) ( ) 6. a. Either talks over students' heads or makes it too simple.
( ) ( ) b. Grading is not objective and impartial.
( ) ( ) c. Has good sense of humor.
( ) ( ) d. Test questions clear and understandable.

---

( ) ( ) 7. a. Assignments not clear, student does not know what is expected of him.
( ) ( ) b. Does not play favorites, impartial in treatment of students.
( ) ( ) c. Enjoys teaching, is enthusiastic and interested in his subject.
( ) ( ) d. Narrow interests, ignores campus life.

---

( ) ( ) 8. a. Chooses a good text.
( ) ( ) b. Course as a whole is well-organized and well-planned.
( ) ( ) c. Test questions vague, not understandable, capable of different interpretations.
( ) ( ) d. Unable to control and guide class discussion.

---

( ) ( ) 9. a. Explanations clear and understandable.
( ) ( ) b. Introduces pertinent supplementary outside material in class.
( ) ( ) c. Presentation is dogmatic, biased, or opinionated.
( ) ( ) d. Shows no interest or enthusiasm in teaching.

---

( ) ( ) 10. a. Emphasizes general principles, central concepts, main trends.
( ) ( ) b. Gives reasonable assignments, neither too long nor too short.
( ) ( ) c. Has poor sense of humor, or none at all.
( ) ( ) d. Plays favorites, treatment of students not impartial.

---

( ) ( ) 11. a. Assignments sporadic and poorly spaced.
( ) ( ) b. Gives proper emphasis to important material.
( ) ( ) c. Has outside interests, takes part in campus activities.
( ) ( ) d. Reads his lectures.

---

( ) ( ) 12. a. Brings irrelevant, unimportant material into class lectures and discussion.
( ) ( ) b. Makes good use of illustrations and examples in lectures.
( ) ( ) c. Presentation is not dogmatic, biased, or opinionated.
( ) ( ) d. Too much emphasis on details in tests.

---

( ) ( ) 13. a. Able to arouse students' interest.
( ) ( ) b. Not helpful with students' problems.
( ) ( ) c. "Rides" or ridicules students.
( ) ( ) d. Shows practical applications of material presented.

---

( ) ( ) 14. a. Able to promote student participation in class.
( ) ( ) b. Assignments too difficult or too easy.
( ) ( ) c. Does not depend solely on text, assigns outside readings.
( ) ( ) d. Poorly poised, not sure of self.

---

Most Least

( ) ( ) 15. a. Expects too much of students, no understanding of their capabilities and level of attainment.
( ) ( ) b. Has no interest in or understanding of his students.
( ) ( ) c. Shows relation of course to other fields.
( ) ( ) d. Tests are "thought-provokers."

---

( ) ( ) 16. a. Is able to put students at ease, establishes informal classroom atmosphere.
( ) ( ) b. Makes no attempt to be friendly with students.
( ) ( ) c. Presents poor personal appearance (dress, neatness, etc.)
( ) ( ) d. Speaking technique good (voice, gestures, etc.)

---

( ) ( ) 17. a. Assigns no supplementary outside material.
( ) ( ) b. Gives students outline and objectives of course.
( ) ( ) c. Has genuine interest in and understanding of his students.
( ) ( ) d. Tests unreasonable and unfair.

---

( ) ( ) 18. a. Does not hand back or discuss tests and papers.
( ) ( ) b. Looks down on students.
( ) ( ) c. Makes certan that students understand material before moving on.
( ) ( ) d. Makes studying enjoyable, students feel like putting forth effort.

---

( ) ( ) 19. a. Lectures well-organized.
( ) ( ) b. Petty strictness.
( ) ( ) c. Poor use of English (grammar, pronunciation, etc.)
( ) ( ) d. Uses techniques designed to make students keep up on work. (Drop quizzes, oral questions, etc.)

---

( ) ( ) 20. a. Explains his grading system to students.
( ) ( ) b. Forgets to call for assigned papers, or does not call for them when due.
( ) ( ) c. Lectures are interesting and stimulating.
( ) ( ) d. Unpleasant, sour disposition.

## Literature Cited

Flanagan, John C. 1939. General Considerations in the Selection of Test Items and a Short Method of Estimating the Product-Moment Coefficient from Data at the Tails of the Distribution. *Journal of Educational Psychology*. 30: 674-680.

Sisson, E. Donald. 1948. Forced Choice — The New Army Rating. *Personnel Psychology*. 1: 365-381.

Department of Psychology
Grinnell College
Grinnell, Iowa