To determine whether placement of items on a teacher rating scale affects the factor structure underlying the scale and to determine whether changing the item format alters the ratings given, a 12-item, high-inference student rating scale was developed containing two global items pertaining to overall teacher effectiveness and 10 evaluative items pertaining to more specific teacher characteristics. Four different formats were constructed and randomly distributed to 363 college students within 27 university classes. The results of a principal components analysis suggest that the factor structure underlying the rating instrument is stable and not affected by changes in the placement of items. The nonsignificant results from a multivariate analysis of variance (MANOVA) provide additional support for the stability of the ratings. The MANOVA results in particular suggest that student ratings do not increase or decrease as a result of changes in item placement. The results of both analyses detract from the view that the halo effect may be influenced by the positioning of items. There is thus little reason to believe that preliminary items may set the stage for responses on following items. (RC)
The Influence of Format Change on the Halo Effect of Student Ratings

Robin D. Froman
University of Connecticut

The Influence of Format Change on the Halo Effect of Student Ratings

Robin D. Froman
University of Connecticut

Introduction and Review

One of the chronic problems of student ratings has been establishing content validity. Studies seeking to establish the technical adequacy of student ratings have generally supported the reliability of the measures, but validity remains controversial. It is not clear whether student ratings of teacher performance are actually reflections of teacher characteristics and teaching ability, or if they are in some part a reflection of student characteristics.

The halo effect has been claimed to cloud the interpretation of rating outcomes. The halo effect, in brief, is the tendency for raters to evaluate conceptually different characteristics as though they all represented the same thing. It is the tendency for prior information, or one impression or characteristic, to influence all other impressions. In other words, if an instructor receives a high rating on one characteristic, it is likely that he will receive a high rating on other characteristics.

Person-perception research has built a considerable argument for the existence of implicit personality theories, one possible cause or explanation for the halo effect. In a review of the research on the effects of implicit personality theory, Schneider (1973) suggests that
Raters: Rating instruments are more affected by the preconceived personality theory of the rater than by the characteristics of the person being rated. Evidence is cited in the review to illustrate that similar trait structures emerge when different rating instruments are factor analyzed. These structures are virtually unaffected by the stimulus object, the familiarity of the rater with the ratee, differences in samples, or differences in situations (see, for example, Magoon and Price, 1972).

Within the literature pertaining specifically to teacher evaluation instruments, consideration has been given to the effect of implicit personality theory. Johnson, Rhodes, and Rumery (1973) suggest that teacher rating instruments do not necessarily furnish information about teacher performance, but rather, provide information about the acceptability of teacher characteristics and performance. Thus, teacher rating scales would appear to provide more information about student perceptions and implicit personality theories than about actual teacher behaviors. This notion is supported by the results of a study by Magoon and Price (1972). When comparing the results of principal component analyses of 21 sets of rating data between and within classrooms, the authors found that the major dimensions which emerged were very similar in structure. These researchers concluded that these striking similarities provided evidence that ratings reveal more about student preconceptions about the relationships between teacher traits than about the quality of instruction. In a later investigation of this possibility, Bausell and Magoon (1972) found a correlation of .67 between evaluations of teaching performance made on the first day of class and evaluations made at the end of the semester. In addition, the authors found that the changes in ratings
that did occur were predictable and related to changes in the grade a
student expected to receive. If implicit personality theories are oper-
atating, one wonders whether stability estimates (of reliability) indicate
consistency in teacher performance or consistency in students' beliefs.

Several studies have been conducted to determine what variables do
or do not affect the bias in ratings which result from a rater's implicit
personality theory. Schneider (1973) reported that occupation of the
rater, caffeine arousal, and the perceived importance of the evaluation
were found to influence the factor structures which emerged from analyses
of rating instruments. Slater and Owen (1974) found that the major area
of study of students affected the factor structure which emerged when
students' ratings of hypothetical "ideal professor" were analyzed. The
differences in the factors due to a student's major suggest that one's
field of study may be a relevant predictor of the bias or implicit per-
sonality theory a student has.

Dixon and Dixon (1975) found this bias to be affected by the over-
all perception a student has of a teacher's competence. A comparison
of teachers rated as "exceptionally good" versus teachers rated as
"exceptionally poor" yielded two different sets of factor structures.
The differences in the factor structures imply that students have dif-
ferent implicit personality theories with respect to good teachers and
poor teachers.

A study by Follman, Lucoff, Small, and Power (1974) reports quanti-
tative differences in ratings given to teachers as a result of format
changes in scaling, coding, and response alternatives. There were,
however, no analyses reported which would determine if there were changes
in the factor structures as a result of format changes. Thus the Follman et al. study shows evidence of changes in the mean ratings awarded based upon format changes, but does not determine if the dimensions underlying student bias change as a result of format changes.

The purpose of the present investigation was to determine if simple format changes would alter the existing halo effect and hence influence student ratings of teacher performance. The research was designed to find out whether the placement of questions on a rating scale is responsible for any halo effect. In particular, it was felt that global items pertaining to teacher effectiveness might provide a certain "tone" or "set" for students as they respond to subsequent questions. That is, having responded to initial items, the students might simply "follow suit" on the rest of the items on a scale.

The specific objectives of this study were twofold:

1. To determine whether placement of items on a teacher rating scale affects the factor structure underlying the scale; and

2. to determine whether changing the item format alters the ratings given.

Procedure

A twelve-item, high-inference student rating scale was developed containing two global items pertaining to overall teacher effectiveness and ten evaluative items pertaining to more specific teacher characteristics (see Appendix A).¹

Four different formats were constructed, varying the items as follows:

1. The two global items were at the top of the rating instrument;

¹For more information concerning the construction, validity, and reliability of the rating scale refer to, Student Ratings; What is the Frame of Reference? by A. Lolli and S. Owen, a paper presented at the Annual Convention of NCME, San Francisco, 1976.
2. The two global items were at the bottom of the rating instrument;
3. The two global items were absent from the rating instrument;
4. The two global items were the only items present on the form.

The four formats are shown in Appendices A through D. The four forms were randomly distributed to 363 students within 27 university classes. Undergraduate and graduate students from four academic areas were represented in the sample.

Data from all classes were pooled and three principal components analyses were conducted to identify the factor structures underlying the ten evaluative items. Data from forms 1, 2, and 3 were considered in the factor analysis. A one-way MANOVA was applied to the data from forms 1, 2, and 4, considering the two global items as the dependent variables. A second one-way MANOVA was performed on forms 1, 2, and 3, considering the ten evaluative items as the dependent measures (see Figure 1).

Results

Table 1 shows the results of the principal components analysis. The factor structures across the three forms containing the ten evaluative items were very similar. The strongest five items loading above .40 on Factor I were the same across the three forms, with slight changes in the order of item importance. The average percent of variance accounted for by Factor I was .39; the range from .34 to .42. Analyses of the data from form 3 yielded only two factors; the data from forms...

---

2 The academic areas represented in the sample were education, psychology, engineering and history.
1 and 2 yielded three factors. The two factor solution resulting from form 3 was apparently the consequence of collapsing the second and third factors which emerged on the other two forms into one factor. The collapsing of the factors was probably due to the eigenvalue criterion of 1.0. A third factor, quite similar to the third factor seen in forms 1 and 2, developed an eigenvalue of .95 and was not retained for rotation.

The slight changes in the relative importance of items in defining factors two and three on forms 1 and 2 can be explained as the result of a restriction of range on those items. In particular, items 5 and 7 had average variances of .37 and .21 respectively across the three forms. The average variances of the eight other evaluative items across the three forms compared in the factor analysis was .76.

---

Insert Table 1 about here

---

Two MANOVA's were conducted to determine if format changes would alter the ratings given. Neither MANOVA showed differences in the mean vectors of the dependent variables. The MANOVA analyzing the mean vectors of the ten evaluative items produced an $F$ value of 0.83, $p < .68$, with degrees of freedom of 20 and 702. The analysis of the forms containing only the two global items showed an $F$ value of .78, $p < .54$, with degrees of freedom of 4 and 710.

Conclusions

The results of the principal components analyses suggest that the factor structure underlying the rating instrument is stable and not affected by changes in the placement of the items. The nonsignificant
results from the MANOVA provide additional support for the stability of the ratings. The MANOVA results in particular suggest that student ratings do not increase or decrease as a result of changes in item placement. The results of both analyses detract from the view that the halo effect may be influenced by the positioning of items. There is thus little reason to believe that preliminary items may "set the stage" for responses on following items.

Although this research has not pinpointed the reasons for the halo effect, the negative findings may be viewed as a fortunate outcome. Researchers in instrument development may be somewhat more confident that general format revisions will not influence the technical adequacy of teacher rating scales.
APPENDIX A (Form 1)
The University of Connecticut Survey of Courses and Teaching

<table>
<thead>
<tr>
<th>Department</th>
<th>Course #</th>
<th>Instructor</th>
</tr>
</thead>
</table>

**DIRECTIONS:** Student ratings can be a powerful type of feedback to the instructor. Please be as objective as you can. **DO NOT SIGN YOUR NAME.**

**Your Semester:**

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>Grad.</th>
</tr>
</thead>
</table>

Is this course in your major field? [ ] Yes [ ] No

**Expected grade in course:**

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>F</th>
<th>PASS</th>
</tr>
</thead>
</table>

**Cumulative Grade Average:**

<table>
<thead>
<tr>
<th>1.0-1.5</th>
<th>1.5-2.0</th>
<th>2.0-2.5</th>
<th>2.5-3.0</th>
<th>3.0-3.5</th>
<th>3.5-4.0</th>
</tr>
</thead>
</table>

**Percent of class attendance:**

<table>
<thead>
<tr>
<th>70-100%</th>
<th>60-69%</th>
<th>50-59%</th>
<th>40-49%</th>
<th>30-39%</th>
<th>20-29%</th>
<th>10-19%</th>
</tr>
</thead>
</table>

**Time spent on this course compared to other courses of equal credit:**

<table>
<thead>
<tr>
<th>considerably more</th>
<th>somewhat more</th>
<th>about the same</th>
<th>somewhat less</th>
<th>considerably less</th>
</tr>
</thead>
</table>

**How much have you learned from this course?**

<table>
<thead>
<tr>
<th>almost nothing</th>
<th>a little</th>
<th>a moderate</th>
<th>quite a bit</th>
<th>a great deal</th>
<th>amount</th>
</tr>
</thead>
</table>

**How would you rate this instructor in general, all-around teaching ability?**

<table>
<thead>
<tr>
<th>poor</th>
<th>less than average</th>
<th>average</th>
<th>good</th>
<th>excellent</th>
</tr>
</thead>
</table>

**For the following ten questions, rate each item in two ways:**

a. give your actual rating of the instructor; then
b. mark the second scale according to how you would prefer things to be. Of course, if you are satisfied, the second scale will be marked the same as the first.

**From my observations, this instructor:**

1. presents material in a clear & effective manner
   (preferred) [ ] [ ] [ ] [ ] [ ]

2. stimulates interest
   (preferred) [ ] [ ] [ ] [ ] [ ]

3. makes work assignments & student responsibilities clear
   (preferred) [ ] [ ] [ ] [ ] [ ]

4. uses exam items which stress important aspects of the course
   (preferred) [ ] [ ] [ ] [ ] [ ]

5. grades fairly and impartially
   (preferred) [ ] [ ] [ ] [ ] [ ]

6. demonstrates overall organization
   (preferred) [ ] [ ] [ ] [ ] [ ]

7. meets class regularly and on time
   (preferred) [ ] [ ] [ ] [ ] [ ]

8. fulfills class objectives and obligations
   (preferred) [ ] [ ] [ ] [ ] [ ]

9. is accessible to students both in and out of class
   (preferred) [ ] [ ] [ ] [ ] [ ]

10. shows an interest in and concern for students
    (preferred) [ ] [ ] [ ] [ ] [ ]
APPENDIX B (Form 2)
The University of Connecticut Survey of Courses and Teaching

Department ___________________ Course # ___________ Instructor ___________

Directions: Your instructors will not see these ratings until after grades are awarded. Please be as objective as you can, and DO NOT SIGN YOUR NAME.

Your semester: [ ] 1 2 3 4 5 6 7 8 Grad. Is this course in your major field? [ ] yes [ ] no

Expected grade in course: [ ] A [ ] B [ ] C [ ] D [ ] F [ ] Pass

Cumulative grade average: [ ] 1.0-1.5 [ ] 1.5-2.0 [ ] 2.0-2.5 [ ] 2.5-3.0 [ ] 3.0-3.5 [ ] 3.5-4.0

Percent of class attendance:

<table>
<thead>
<tr>
<th>90-100%</th>
<th>75-89</th>
<th>50-74</th>
<th>25-49</th>
<th>0-24</th>
</tr>
</thead>
</table>

Time spent on this course compared to other courses of equal credit:

[ ] considerably more [ ] somewhat more [ ] about the same [ ] somewhat less [ ] considerably less

For the following ten questions, rate each item in two ways:
a. give your actual rating of the instructor; then
b. mark the second scale according to how you would prefer things to be. Of course, if you're satisfied, the 2nd scale will be marked the same as the 1st.

From my observations, this instructor...

1. presents material in a clear & effective manner (preferred)

2. stimulates interest (preferred)

3. makes work assignments & student responsibilities clear (preferred)

4. uses exam items which stress important aspects of the course (preferred)

5. grades fairly and impartially (preferred)

6. demonstrates overall organization (preferred)

7. meets class regularly and on time (preferred)

8. fulfills class objectives and obligations (preferred)

9. is accessible to students both in and out of class (preferred)

10. shows an interest in and concern for students (preferred)

How much have you learned from this course?

[ ] almost nothing [ ] a little [ ] a moderate amount [ ] quite a bit [ ] a great deal

How would you rate this instructor in general, overall teaching ability?

[ ] poor [ ] less than average [ ] adequate [ ] good [ ] excellent
APPENDIX C (Form 3)
The University of Connecticut Survey of Courses and Teaching

Department ____________________ Course # ____________________ Instructor ____________________

Directions: Your instructors will not see these ratings until after grades are awarded. Please be as objective as you can, and DO NOT SIGN YOUR NAME.

Your semester: 1 2 3 4 5 6 7 8 Grad. Is this course in your major field? [ ] yes [ ] no

Expected grade in course: [ ] A [ ] B [ ] C [ ] D [ ] F [ ] Pass

Cumulative grade average:
- 1.0-1.5
- 1.5-2.0
- 2.0-2.5
- 2.5-3.0
- 3.0-3.5
- 3.5-4.0

Percent of class attendance:
- 90-100%
- 75-89
- 50-74
- 25-49
- 0-24

Time spent on this course compared to other courses of equal credit:
- considerably more
- somewhat more
- about the same
- somewhat less
- considerably less

For the following ten questions, rate each item in two ways:

a. give your actual rating of the instructor; then
b. mark the second scale according to how you would prefer things to be. Of course, if you're satisfied, the second scale will be marked the same as the first.

From my observations, this instructor...

1. presents material in a clear & effective manner... [ ] preferred [ ]
2. stimulates interest... [ ] preferred [ ]
3. makes work assignments & student responsibilities clear... [ ] preferred [ ]
4. uses exam items which stress important aspects of the course... [ ] preferred [ ]
5. grades fairly and impartially... [ ] preferred [ ]
6. demonstrates overall organization... [ ] preferred [ ]
7. meets class regularly and on time... [ ] preferred [ ]
8. fulfills class objectives and obligations... [ ] preferred [ ]
9. is accessible to students both in and out of class... [ ] preferred [ ]
10. shows an interest in and concern for students... [ ] preferred [ ]
APPENDIX D (Form 4)
The University of Connecticut Survey of Courses and Teaching

Department ____________________  Course # __________  Instructor ____________________

Directions: Your instructor will not see these ratings until after grades are awarded. Please be as objective as you can, and DO NOT SIGN YOUR NAME.

Your semester: 1 2 3 4 5 6 7 8 Grad. Is this course in your major field? 
yes  no

Expected grade in course: A B C D F Pass  Cumulative grade average:
1.0-1.5
1.5-2.0
2.0-2.5
2.5-3.0
3.0-3.5
3.5-4.0

Percent of class attendance:
90-100%
75-89
60-74
50-64
25-49
0-24

Time spent on this course compared to other courses of equal credit:
considerably more
somewhat more
about the same
somewhat less
considerably less

How much have you learned from this course?
almost nothing
a little
a moderate
quite a bit
a great deal

How would you rate this instructor in general, overall teaching ability?
poor
less than average
adequate
average
good
excellent

14
BIBLIOGRAPHY


FIGURE 1
Description of MANOVA's to Determine if Format Changes Alter Ratings Awarded

<table>
<thead>
<tr>
<th>Form 1</th>
<th>Form 2</th>
<th>Form 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two global items on top</td>
<td>Two global items on bottom</td>
<td>Two global items only,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>evaluative items absent</td>
</tr>
<tr>
<td>Form 1</td>
<td>Form 2</td>
<td></td>
</tr>
<tr>
<td>Ten evaluative items on</td>
<td>Ten evaluative items on</td>
<td>Ten evaluative items only,</td>
</tr>
<tr>
<td>bottom</td>
<td>top</td>
<td>global absent</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TABLE 1
Summary of Three Principal Components Analyses of Ten Items Contained on Forms 1, 2, and 3 of the Teacher Rating Scale

<table>
<thead>
<tr>
<th>Item</th>
<th>Loading</th>
<th>Item</th>
<th>Loading</th>
<th>Item</th>
<th>Loading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Form 1</td>
<td>Form 2</td>
<td>Form 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>.808</td>
<td>1.</td>
<td>.892</td>
<td>1.</td>
<td>.911</td>
</tr>
<tr>
<td>1.</td>
<td>.801</td>
<td>2.</td>
<td>.848</td>
<td>2.</td>
<td>.866</td>
</tr>
<tr>
<td>8.</td>
<td>.786</td>
<td>8.</td>
<td>.842</td>
<td>6.</td>
<td>.816</td>
</tr>
<tr>
<td>6.</td>
<td>.785</td>
<td>6.</td>
<td>.836</td>
<td>8.</td>
<td>.647</td>
</tr>
<tr>
<td>10.</td>
<td>.638</td>
<td>10.</td>
<td>.412</td>
<td>10.</td>
<td>.507</td>
</tr>
<tr>
<td>3.</td>
<td>.468</td>
<td>3.</td>
<td>.405</td>
<td>5.</td>
<td>.471</td>
</tr>
</tbody>
</table>

Factor I

<table>
<thead>
<tr>
<th>Item</th>
<th>Loading</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.</td>
<td>.744</td>
</tr>
<tr>
<td>4.</td>
<td>.714</td>
</tr>
<tr>
<td>3.</td>
<td>.400</td>
</tr>
</tbody>
</table>

Factor II

<table>
<thead>
<tr>
<th>Item</th>
<th>Loading</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.</td>
<td>.810</td>
</tr>
<tr>
<td>7.</td>
<td>.704</td>
</tr>
</tbody>
</table>

Factor III

17
# TABLE 2

Means and Standard Deviations of the Four Formats of the Teacher Rating Scale

<table>
<thead>
<tr>
<th>Item</th>
<th>Form 1 N=120</th>
<th></th>
<th>Form 2 N=118</th>
<th></th>
<th>Form 3 N=125</th>
<th></th>
<th>Form 4 N=121</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\bar{X}$</td>
<td>sd</td>
<td>$\bar{X}$</td>
<td>sd</td>
<td>$\bar{X}$</td>
<td>sd</td>
<td>$\bar{X}$</td>
</tr>
<tr>
<td>1.</td>
<td>4.27</td>
<td>.80</td>
<td>4.37</td>
<td>.88</td>
<td>4.26</td>
<td>.84</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>4.22</td>
<td>.94</td>
<td>4.16</td>
<td>.95</td>
<td>4.11</td>
<td>.94</td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>4.41</td>
<td>.93</td>
<td>4.39</td>
<td>.86</td>
<td>4.45</td>
<td>.87</td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>4.38</td>
<td>1.10</td>
<td>4.02</td>
<td>1.23</td>
<td>4.15</td>
<td>.97</td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>4.74</td>
<td>.49</td>
<td>4.60</td>
<td>.63</td>
<td>4.65</td>
<td>.69</td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>4.26</td>
<td>.91</td>
<td>4.19</td>
<td>1.05</td>
<td>4.16</td>
<td>.92</td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>4.78</td>
<td>.55</td>
<td>4.82</td>
<td>.43</td>
<td>4.86</td>
<td>.37</td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td>4.47</td>
<td>.66</td>
<td>4.47</td>
<td>.82</td>
<td>4.43</td>
<td>.80</td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td>4.62</td>
<td>.66</td>
<td>4.58</td>
<td>.74</td>
<td>4.54</td>
<td>.65</td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td>4.63</td>
<td>.59</td>
<td>4.58</td>
<td>.77</td>
<td>4.57</td>
<td>.75</td>
<td></td>
</tr>
<tr>
<td>Global 1</td>
<td>4.04</td>
<td>.96</td>
<td>3.91</td>
<td>1.00</td>
<td>3.99</td>
<td>.84</td>
<td></td>
</tr>
<tr>
<td>Global 2</td>
<td>4.24</td>
<td>.76</td>
<td>4.27</td>
<td>.87</td>
<td>4.29</td>
<td>.83</td>
<td></td>
</tr>
</tbody>
</table>

18