THE QUANTITATIVE ASSESSMENT OF HALO EFFECT

APPROVED:

[Signatures]

Major Professor

Minor Professor

Dean of the School of Education

Dean of the Graduate School
THE QUANTITATIVE ASSESSMENT OF HALO EFFECT

THESIS

Presented to the Graduate Council of the North Texas State University in Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

By

Walter William Gilchrist, B. S.

Denton, Texas

# TABLE OF CONTENTS

LIST OF TABLES ........................................ iv

Chapter

I. INTRODUCTION .......................................... 1

   Theoretical Background
   Purpose of the Study
   Related Literature

II. PROCEDURE ............................................ 7

   Subjects
   Methods for Collecting Data
   Statistical Treatment

III. RESULTS ............................................. 11

IV. SUMMARY .............................................. 19

   Recommendations

BIBLIOGRAPHY ........................................... 22
LIST OF TABLES

Table                                                                 Page

I. Intercorrelation of Sociometric and Objective Measures ............... 12
CHAPTER I

INTRODUCTION

Theoretical Background

For fifty years halo has been a central problem in the theory of absolute judgment. It is now common knowledge that persons who are liked by others may possess certain personality traits, verbal and physical habits, and other characteristics which when noticed in an individual who is not liked, only tend to intensify this dislike of them. These same traits when found in someone that is liked may be accepted, overlooked, or forgotten. When a certain person is liked, it is the tendency to incorporate halo into the judgment of this person's skill, ability, and knowledge in many other areas. That is, this person is often given credit where credit is not truly in order or due. Whether a person is liked or disliked has been considered to be a major determinant of how he is rated in areas necessitating skill, ability, trustworthiness, and other such traits. Halo effect is considered to be the result of a tendency for a rate to base ratings of all traits on a general impression formed by consideration of only a few traits.
Ratings by one's peers are widely used both for research and administrative purposes. This is especially true in the Armed Forces in which all branches of the service make use of these ratings for the identification and prediction of leadership ability. Efforts to reduce the over or under evaluation of individuals in these judgments could take two forms. One of these is by proper training of the raters to increase the objectivity of their appraisals. The other would be to employ some mathematical means for correcting the ratings or choices. The former is often time consuming and not always feasible.

Peer ratings, in general, have been popular due to three characteristics. They are easy to obtain in almost any organized group, reliability is usually satisfactory and often high, and they will usually correlate higher than test scores and other variables with most criteria.¹ If it is possible to eliminate or remove halo effect from such ratings a more accurate and informative score would remain.

Purpose of the Study

Peer ratings have at least three sources of variance. These are true score variance, halo effect, and random error. It is the purpose of this investigation to quantitatively

assess the effect of halo in peer ratings of individuals in the areas of friendship, mental ability, and level of anxiety, by identifying the discrepancy between these ratings and the measurement of these characteristics by standardized tests.

Within the above framework, the following hypotheses will be examined:

Hypothesis 1. Halo effect would appear in the peer ratings in all areas.

Hypothesis 2. Peer group ratings of high and low mental ability will be considerably more accurate than the peer group ratings of anxiety. Therefore, more halo will be found in the anxiety ratings.

Hypothesis 3. When halo is removed from the peer group ratings, the remainder will correlate more highly with the objective measurements of the same criteria.

Related Literature

In 1920 E. L. Thorndike observed that the intercorrelations between ratings for various traits on army officers were usually high and uniform. Thorndike reasoned that this might indicate several things; either the traits on which these men were rated were very similar, or the raters were in error. The former seemed unlikely and therefore Thorndike concluded that the fault must be with the judges. Either their opinion concerning the first trait colored all subsequent
traits or they were unable to discriminate between the different traits on which they were to rate others. The former seemed more likely. Thorndike, however, was certain that a constant error was present in the judgment of the raters which led them to rate an individual about the same on all traits to be considered. He referred to this condition as the "constant error of the halo." Later writers changed this to "halo effect" or "halo tendency." Thorndike recognized that the constant error may be in either direction, negative or positive.

Borrowing from Thorndike, halo effect is regarded in this thesis as the tendency to rate an individual on traits two, three, and four high or low because one rated him high or low on trait one. The rater's error or halo lies in his tendency to continue with the same judgment regardless of the trait considered.

For the most part, ways of reducing halo in the original material have been attempted. Such means as training the raters and instructing the raters to rate all persons on one criterion or trait before proceeding to the second trait, have been employed in order to minimize the halo effect.

---


Few attempts have been made to statistically remove halo from the ratings per se. One such attempt to statistically remove halo was performed by Grant through the use of factor analysis technique. Grant found an average correlation of .55 between ratings made by 110 supervisors in a large insurance company. In this study there were eighteen characteristics to be rated. A considerable portion of the variance of these ratings was due to halo.\(^5\)

Mayo found a correlation of .66 between ratings of effort and intelligence.\(^6\) Since there is no reason to believe that the most intelligent individuals always put out the greatest effort, or vice versa, it seems likely that this high relationship is primarily due to halo.

In a study conducted by Trites, Kubala, and Cobb at the School of Aviation Medicine, Randolph AFB, Texas, 337 subjects rated each other on ten characteristics. The first rating scale consisted of four characteristics and the second scale consisted of six characteristics. Both scales asked the subjects to rate their peers on the characteristic of cooperativeness. The first scale required ratings on "team cooperativeness" and defined this in the following statement. "In selecting a team for some sport, imagine all men in your

---


squad have exactly the same ability. Your selection should be based on the tendency of each man to cooperate with other team members, to work with the team as a unit, and to subordinate himself to the goals and the operation of the team as a whole. You want men on your team who could do this. The second scale asked the subjects to rate their peers on "cooperation." Cooperation was defined "as the ability and willingness to work in harmony for and with others." The correlation between these two should be higher than the other more independent characteristics due to the great similarity. However, the correlation between the two ratings of cooperativeness was .68. The mean inter-correlation of all ten ratings was also .68. These data indicate that the ratings of the characteristics are not really independent. Apparently the raters did not distinguish between the characteristics, but responded to each in terms of a generalized tendency to perceive the person in a particular way.

---


8 Ibid., p. 11.
CHAPTER II

PROCEDURE

Subjects

The subjects used in this study of halo effect were thirty-eight high school students. They were the entire graduating senior class of a small high school in Denton County, composed of twenty-three females and fifteen males. The group was selected because most of the class had been together for more than six years and none for less than three years. It was considered important that the subjects know each other fairly well in order to make realistic choices. This class was considered to be a typical high school senior class. The mean intelligence quotient as determined by the Otis Quick Scoring Mental Ability test was 103.7. The mean age was eighteen years and two months.

Methods for Collecting Data

Peer group ratings were gathered on three criteria: like-dislike, anxiety, and mental ability. Each person was given three class rolls. The following instructions were given.

Everyone should have three sheets of paper on his desk. These are three identical lists of names of the seniors in alphabetical order.
Put your name in the upper right-hand corner of all three sheets.

Take one of these sheets and put the other two aside. In the one you have selected cross out the words "Taylor Scale" and "Mental Ability."

In the box to the right of the name put a plus mark after the names of the five persons you like the most.

On this same sheet, put a minus sign after the five persons' names you like the least.

When you have finished put the sheet face down on your desk.

Now, take another of the sheets with the list of your classmates' names on it. On this one please cross out the words "Sociometric" and "Mental Ability." Put a plus mark to the right of the names of the five persons you believe to have the least anxiety. Here is a definition of anxiety: Anxiety is a common experience we have all felt in varying degrees. We feel perhaps something bad is going to happen to us. Or, we may feel disorganized. We dread something and yet do not know just exactly what it is we dread. In an anxiety state we have somewhat the same feelings as we do in a state of fear, yet we do not know exactly what it is that arouses these feeling within us. Stage fright and the feeling you have before a forthcoming test are examples of anxiety.

Now put the minus sign after the names of those five persons you believe to have the most anxiety.
When you have finished please lay the sheet face down on your desk.

On the last list of names please cross out the words "Taylor Scale" and Sociometric." Put a plus mark after the names of the five persons you believe to have the highest mental ability or I. Q.

Now, put a minus sign after the names of the five persons you believe to have the least mental ability or I. Q.

When finished, put this sheet face down on the stack with the other sheets and they will be picked up.

The objective measurement of anxiety was obtained from scores on the Taylor Manifest Anxiety Scale, and those of mental ability were obtained from the Otis Quick Scoring Mental Ability Test. These were administered to the group after a brief rest period.

All three of the previously mentioned traits on the peer rating scales were scored in the same way. Each student was given a plus one for each positive choice on the trait. Each student was given a minus one for each negative choice on the trait. The score was the algebraic sum of the thirty-seven possible choices each person received. The theoretical maximum range of scores was from a plus thirty-seven to a minus thirty-seven, as each person was rated by the other thirty-seven persons in the class.
Statistical Treatment

The Pearson-Product-Moment method of correlation was used in determining the degree of relationship between peer ratings and the objective ratings of friendship, mental ability, and anxiety. Intercorrelational indices were obtained for each of the three areas with every other area.

CHAPTER III

RESULTS

In order to determine the degree of halo present in peer ratings, it was first necessary to correlate the subjective choices with the objective measures of anxiety and mental ability. Positive and negative peer group choices on the sociometric portion were separately correlated with the scores obtained on the Taylor Anxiety Scale and the Otis Quick Scoring Mental Ability Test. The sum of the positive and negative choices were all correlated with the objective measures. In order to work with only positive numbers the largest negative sum was added to all other sums of positive and negative choices. Hereafter the term "sum" will be used to designate the positive choices plus the negative choices plus thirty. Correlations of positive, negative and sum were performed in order to see which of these measurements appeared more closely correlated with the objective measurements.

The results of the statistical analysis of the data are presented in Table I. As seen in Table I, the degree of relationship between positive anxiety peer group choices and scores received on the Taylor Anxiety Scale resulted in a correlation of -.07. The negative choices of anxiety when
TABLE I
INTERCORRELATION OF SOCIOMETRIC
AND OBJECTIVE MEASURES

<table>
<thead>
<tr>
<th>Measures</th>
<th>Sociometric Sums</th>
<th>Anxiety Sums</th>
<th>Mental Ability Sums</th>
<th>Taylor Anxiety Scores</th>
<th>Otis Mental Ability Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sociometric Sums</td>
<td>X</td>
<td>.10</td>
<td>.34</td>
<td>.02</td>
<td>.28</td>
</tr>
<tr>
<td>Anxiety Sums</td>
<td></td>
<td>X</td>
<td>-.08</td>
<td>-.11</td>
<td>.25</td>
</tr>
<tr>
<td>Mental Ability Sums</td>
<td></td>
<td></td>
<td>X</td>
<td>.07</td>
<td>.71</td>
</tr>
<tr>
<td>Taylor Anxiety Scores</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>.18</td>
</tr>
<tr>
<td>Otis Mental Ability Scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

correlated with the scores obtained on the Taylor Scale resulted in a correlation of .12. The sum of the peer group choices for anxiety when correlated with the scores on the Taylor Scale yielded a correlation of -.11. None of these are significant. The positive peer group choices on mental ability were correlated with the scores obtained on the Otis Mental Ability Test. This resulted in a correlation of .54. A correlation of the negative choices on the mental ability portion with the scores obtained on the Otis Mental Ability
Test revealed a correlation of \(-0.62\). When the sum of the mental ability portion was correlated with the scores received on the Otis a correlation of \(0.71\) was obtained. These three correlations were all significant at the 1 per cent level of significance. However, the correlations of the sum of choices received on a trait, when correlated with the scores of the objectively measured trait, gave a correlation which was as good as or possibly better than those obtained when the positive or negative choices were correlated individually with the objective scores. This in itself is justification for using the sum of choices rather than positive or negative choices in future correlations. The sum of peer group choices takes into account both positive and negative choices and thereby increases the range, and theoretically increases the reliability.

When like-dislike sums were correlated with anxiety sums the resulting correlation should be primarily due to halo. The magnitude of this correlation should indicate the amount of halo. In this research a correlation of only \(0.10\) was obtained, indicating that insignificant halo was present; however, the correlation was in the expected direction.

A correlation of like-dislike sums with the sums on peer ratings of mental ability should also yield a figure which is primarily a measure of halo. An examination of
this relationship yielded a correlation of .34, indicating the possibility of at least a marginal halo effect. However, since the correlation between like-dislike sums and the Otis Mental Ability Test was .28, the relationship between the rating measure and the objective test makes it appear that intelligence is probably a factor preference. Consequently, the correlation between the like-dislike rating and mental ability rating is probably nothing more than a reflection of this relationship, and not due to halo. Regardless, it was immediately obvious that partial correlation methods could not be used to "purify" the peer ratings of mental ability.

Surprisingly little halo appeared in the choices of this particular group. Since halo has been found so frequently in similar ratings by other groups, the decision was made to investigate the reliability of the ratings obtained in the present study. The subjects were divided into groups for the purpose of computing split half reliabilities by selecting every other name from an alphabetical roster for the first group and putting all remaining names into a second group. Sums were computed for each half and these were correlated.

The analysis of data indicated the lack of halo in the present group was not due to the unreliability of the peer ratings. The split half reliability coefficient for mental ability was .97. The split half reliability coefficient for sociometric, like-dislike, was .74. The split half reliability
coefficient for the anxiety choices was .71. These reliability coefficients are uncorrected. If these coefficients are corrected by the Spearman-Brown formula they are approximately equal in magnitude to those reported by Hollander. ⁹

Even though the over-all reliability was sufficient, there was a possibility that the results may be due to the fact that boys and girls were rating their classmates from a different frame of reference. To check this source of variation a score was obtained for each of the thirty-eight subjects for the number of choices each received from boys and the number of choices each received from girls. It was expected that like-dislike choices which are based for the most part on same sex friendships to be the least related. A correlation of choices given by boys with choices given by girls yielded a coefficient of .32. A correlation of choices on anxiety given by boys with choices given by girls yielded a coefficient of .63. These data suggest that both sexes were making their choices using a similar frame of reference.

A correlation of the like-dislike choices made by boys with the anxiety choices made by boys yielded a coefficient of .21. A correlation of the like-dislike choices made by girls with anxiety choices made by girls yielded a coefficient

Neither of these two coefficients is significant. These two correlations indicated that the lack of halo could not be attributed to using a group composed of both sexes.

As mentioned in the "Related Literature" section of Chapter I, the same method of obtaining ratings has been employed with groups in which the experimental population was military personnel. In these studies halo was consistently found. One possible way to account for the major differences between the amount of halo found in the military population and that of the group used in this study, is the length of time the persons had known each other. The majority of the military persons used in these studies had known each other for around one to four months, whereas the group of high school students used in the present study had known each other from three to six years. This longer period of close association would tend to crystalize one's knowledge of his peers and could result in more realistic ratings, and consequently, less halo.

Hypothesis 2 stated that the peer group ratings of high and low mental ability will be considerably more accurate than the peer group ratings of anxiety. This hypothesis was substantiated by the data. The correlation between the peer ratings of mental ability and the objective measure of mental ability was .71, while the correlation between the peer ratings of anxiety and the objective measure of anxiety was only -.11.
Hence, the results show that mental ability can be much more accurately rated than can anxiety. The corollary of this hypothesis was that more halo would be found in the anxiety ratings. The data did not provide any evidence to either substantiate or refute this corollary because no halo was discernible in either rating.

The lack of a significant correlation between ratings of anxiety and scores received on the Taylor scale could be interpreted in several ways. Perhaps the Taylor Anxiety Scale and the results of the peer group choices on the anxiety portion were evaluating and measuring different aspects of anxiety.

It is interesting to note that the mean score on the Taylor for the group was approximately two and one half times larger than the mean score according to national norms. One way of accounting for this difference may be found in the fact that the experimental group were all graduating high school seniors. It was now that each student had to make certain important decisions such as: whether a college education would be attempted; or where should one look for employment; or should one now contemplate marriage. The making of these decisions and others would establish a rather serious atmosphere. Perhaps the realistic choices made by this group of high school seniors could be attributed to this serious
atmosphere. Perhaps a greater degree of anxiety was present due to the decisions at hand to be made. This could account for the larger number of positive anxiety statements checked on the Taylor Anxiety Scale.
CHAPTER IV

SUMMARY

For many years halo has been the central problem in the theory of absolute judgment. It was the intention of this research to attempt to remove by mathematical means that portion known as halo from the subjective peer group ratings and thereby provide more accurate evaluations. These peer group ratings were gathered on characteristics which could be objectively measured by existing tests.

Within the above framework, the following hypotheses were examined:

Hypothesis 1. Halo effect would appear in the peer ratings in all areas.

Hypothesis 2. Peer group ratings of high and low mental ability will be considerably more accurate than the peer group ratings of anxiety. Therefore more halo, will be found in the anxiety ratings.

Hypothesis 3. When halo is removed from the peer group ratings, the remainder will correlate more highly with the objective measurements of the same criteria.

The experimental group consisted of thirty-eight graduating high school seniors. The subjects were asked to rate
their peers on three characteristics: like-dislike, anxiety, and mental ability. The like-dislike ratings were considered to be pure halo. Anxiety was objectively measured by the Taylor Manifest Anxiety Scale. Mental ability was objectively measured by the Otis Quick Scoring Mental Ability Test.

Intercorrelations were obtained between the objective measurements and the subjective measurements utilizing the Pearson-Product-Moment method of correlation. None of the correlations between the peer ratings was significant, indicating that no significant halo was operating in these ratings.

Reliability of the peer ratings was investigated utilizing the split half method. The obtained reliabilities were of the same order of magnitude as others reported in the literature.

It was noteworthy, however, that the mean number of choices on the Taylor Manifest Anxiety Scale was two and one half times that of the national norm. This may be due to the fact that the subjects were graduating high school seniors and faced with decisions that generate a higher level of anxiety than would be found under other circumstances.

Recommendations

In this research, it would have been to the project's advantage to have done everything possible in order to produce a maximum halo in order to evaluate the proposed method
for the mathematical removal of halo. Halo effect would probably have been present had the experimental group not known one another for such a great length of time, since a six-year sample of behavior probably affords a realistic rating of a characteristic trait. In order to create conditions conducive to halo, future investigations could allow subjects to make their ratings of all traits and characteristics on the same sheet or roster, since Bonney, cited in Chapter I, felt such condition tended to enhance halo.

The following recommendations, though not related to the mathematical removal of halo from ratings may be of interest. Since halo may be related to the degree of acquaintance of individuals, it may be of value to study variations in degree of halo with groups who have known one another for different periods of time.
BIBLIOGRAPHY

Books

Bonney, Merl E., Mental Health in Education, Boston, Allyn and Bacon, Inc., 1960.


Articles


Reports

Trites, David K., Kubala, Albert L., and Cobb, Bart B., Criterion Dimensions of Adaptability to Pilot Training, School of Aviation Medicine, USAF, Randolph AFB, Texas, 1959.