


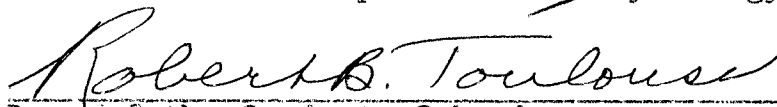
PEER ASSESSMENT OF LEADERSHIP STYLE AND ITS
RELATION TO PRODUCTIVITY

APPROVED:


Major Professor


Minor Professor


Director of the Department of Psychology


Dean of the Graduate School

Day

Swanson, Ronald G., Peer Assessment of Leadership Style and its Relation to Productivity. Master of Science (Psychology), May, 1973, 31 pp., 5 tables, references, 65 titles.

The problem with which this investigation is concerned is that of determining if peer appraisal of leadership style is related to productivity. Previous research in this area has used subordinate evaluation of leadership style, with subjective superior ratings as the performance criterion. The present study uses peers to evaluate style of leadership, with a more objective multiple-input productivity criterion.

The development of the Consideration and Structure dimensions of leadership is included in the introduction, as well as the history of the instrument that measures them. The reliability, stability, and validity of peer ratings are also presented. The weaknesses of subjective criterion measures are examined, and a rationale for a more objective multiple input criterion is developed.

Previous research has shown evidence that superiors rated higher those leaders who were high on the Structure dimension, while subordinates rated higher those leaders who were high on the Consideration dimension. Since peers are neither superiors nor subordinates, it was expected that they would see value in being high on either dimension.

In light of the research findings presented, three hypotheses are put forth for testing. Hypothesis 1 proposes that those leaders rated by their peers as being

high on the Consideration dimension will have significantly higher productivity scores than those rated as being low on Consideration. Hypothesis 2 proposes that those leaders rated by their peers as being high on the Structure dimension will have significantly higher productivity scores than those rated as being low on Structure. Hypothesis 3 proposes that those leaders rated by their peers as being high on both dimensions will have significantly higher productivity scores than any other group of leaders.

Sixty-one instructor pilots from a U.S. Air Force undergraduate pilot training base are used as subjects. Each of these instructors had a minimum of three Supervisory Behavior Description questionnaires filled out on him by his peers, and each instructor pilot had been in his present organization a minimum of six months.

The productivity measure consists of five separate performance inputs, which include flight instructional scores, flight proficiency scores, written test scores, number of errors committed, and flight commander evaluation rankings. These aspect scores are averaged and/or weighted into one overall measure of performance/productivity.

Hypotheses 1 and 2 are fully supported at the .05 level of significance when significant main effects via a 2×2 analysis of variance are demonstrated. Since no significant interaction was evident, Hypothesis 3 could not be fully supported by any definitive statistical test that would

infer significance. An inspection of the cell means gave superficial support to Hypothesis 3, since the group rated as high on both dimensions did have the highest overall mean productivity score.

This report concludes that both peer ratings and the Supervisory Behavior Description dimensions are useful tools in leadership research. The more objective multiple-input productivity criterion is seen as a step forward in the field, and as a measure in which researchers can place more confidence. Implications for further study are discussed and several limitations of the present study are examined.

PEER ASSESSMENT OF LEADERSHIP STYLE AND ITS
RELATION TO PRODUCTIVITY

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

Ronald G. Swanson, B. S.

Denton, Texas

May, 1973

LIST OF TABLES

Table	Page
1. Weights Assigned Relative to Total Errors Assessed	22
2. Weights Assigned Relative to Flight Commander Ranking	23
3. Sample Calculations of Highest and Lowest Overall Productivity Scores	25
4. Number of Subjects and Mean Productivity by Leadership Style	25
5. Summary of Analysis of Variance	26

PEER ASSESSMENT OF LEADERSHIP STYLE AND ITS
RELATION TO PRODUCTIVITY

The relationship of leadership style to productivity has been a subject of great interest to researchers in a number of fields, including industrial psychology, educational psychology, military psychology, and all forms of administration. A basic problem in studies of this type has been in the area of performance and productivity measurement. Korman (1966) summarized the research on this relationship through 1965 and cited only five studies that used objective measures of performance as criteria. Of these five, only one (Parker, 1963) used group productivity as a type of criteria, with the remainder using such measures as corrected salary (Spitzer & McNamara, 1964), grievances (Fleishman & Harris, 1962), and absenteeism, accidents, and turnover (Fleishman, Harris & Burtt, 1955). The correlations between these productivity measures and style of leadership were inconclusive, with some reaching high significance and others showing no significances. The remainder of studies cited used ratings by superiors, peers, self, and subordinates as criteria. A review of the literature since 1965 conducted by the present author revealed no improvements in measures of productivity whatever.

With the use of a more objective measure of productivity, the relationship of leader behavior to productivity should prove to be more stable and reliable. This was one of the objectives of the present study. The other innovation to be used in this study was peer assessment of leadership style instead of the previously used subordinate ratings. In addition, the productivity of the leader himself was used as a criterion instead of a composite score of those subordinate to him.

Before changing several important variables such as those mentioned and launching into the study itself, it is necessary to first establish the credibility of the measuring instrument used, the validity of peer ratings, and the need for better productivity appraisal.

The instrument used to assess style of leadership was the result of over twenty years of research, most of which was done using military flying crews as subjects. In the late 1940's and early 1950's, a substantial amount of research was done on the dynamics of small groups. This period of time found our nation just ending one war and beginning another. Military strength levels were quite high and personnel were dispersed around the world, making them accessible to researchers in many locations.

One of the first areas to be examined in the group process was the role and function of the leader. Adorno, Frenkel-Brunswick, Levinson, and Sanford (1950) did a study

on the authoritarian personality, and suggested that the authoritarian as a leader would more likely be accepted by his superiors than the equalitarian leader. Based upon this, Rohde (1951) used B-29 bomber crews and tested the proposition that the authoritarian as a follower would be more likely accepted by the leader than the equalitarian. The results indicated that authoritarian people were not well accepted as either leaders or followers. Authoritarian persons were not desired as friends, did not inspire confidence, and were not looked upon as good risks in combat. Hollander (1953) found similar results among naval preflight school cadets. The amount of authoritarianism began to appear as a cogent factor in leader behavior.

One of the largest and most significant research programs on leadership was conducted at Ohio State University during the years 1946-1956. While this program was responsible for a variety of significant findings (Stogdill & Coons, 1957), it is likely that its most important contribution was related to the amount of authoritarianism possessed by a leader, and its effect on the preferences and performance of subordinates.

Initially it was decided that there must be certain dimensions along which a leader's behavior varied, and that these dimensions must be labeled, quantified, and measured. Hemphill and Coons (1957) tentatively designated nine dimensions of leader behavior, which were integration, communication, production emphasis, representation, fraternization,

organization, evaluation, initiation, and domination. A 150-item questionnaire was devised to measure these various dimensions. The questionnaire, called the Leader Behavior Description Questionnaire (LBDQ), used the method of summated ratings with scaled multiple choice responses.

The number of dimensions measured and the number of items on the LBDQ were found to be prohibitive for research where time and economy were at stake. For this reason, Halpin and Winer (1957) reduced the number of LBDQ items to 130, administered the scale to fifty-two B-50 bomber crews, and factor-analyzed the data obtained. The results showed that only two factors, Consideration and Initiating Structure, accounted for 83% of the total variance. Many of the LBDQ items were irrelevant to these two factors and were discarded to make the total questionnaire 80 items in length.

These two scales, Consideration and Initiation, were found to be correlated only to a low degree, and sufficiently independent to permit their use as separate measures of different kinds of behavior.

The Consideration dimension reflected the extent to which an individual was likely to have job relationships characterized by mutual trust, respect for subordinates' ideas, and consideration of their feelings. A high score was indicative of a climate of good rapport and two-way communication. A low score indicated that the supervisor was likely to be impersonal in his relations with group members.

The Initiating Structure dimension reflected the extent to which an individual was likely to define and structure his role and those of his subordinates toward goal attainment. A high score characterized individuals who played a more active role in directing group activities through planning, communicating information, scheduling, trying out ideas, etc.

(Fleishman (1953a, 1953b) continued the study of the LBDQ via more extensive factor analyses. A reorganization of the items into relatively more independent categories of leader behavior was accomplished. Two major factors or dimensions remained once again: Consideration and Initiation. Based on item dimension loadings derived from an industrial population, two revised scoring keys were developed--one for Consideration and one for Initiating Structure. Twenty-eight items best met the criteria for Consideration and twenty items for Initiating Structure were selected. Hence, the questionnaire now consisted of forty-eight items and was renamed the Supervisory Behavior Description (SBD) (Fleishman, 1970). The SBD was tested on an industrial population sample, and the results indicated (via orthogonal loading factors) that the two dimensions were quite independent.

During approximately the same time period, a number of other studies were revealing similar findings. Halpin (1954) used the LBDQ on B-29 bomber crews and found that the crew's satisfaction with its commander was positively correlated with the Consideration dimension scores and negatively related

to the Initiating Structure scores. However, the superior's ratings of the commanders were negatively related to the Consideration scores and positively related to the Initiating Structure scores. Thus, superiors and subordinates perceive the two dimensions in opposite ways, with subordinates preferring consideration in their leader, and superiors preferring task orientation in their commanders. This difference confronts the leader with conflicting role expectations and constitutes what the author calls "the dilemma of leadership."

Halpin (1957) performed a similar study which again showed that wing and squadron superiors rate favorably the performance of those commanders who show high Initiating Structure behavior. The crews continued to prefer leaders who were high in Consideration behavior. The critical finding in this study was that if leaders must satisfy both superiors and subordinates, who demand very different modes of behavior, we do best by choosing leaders who are above average on both behavior dimensions.

In an attempt to broaden the scope and use of the LBDQ, Halpin (1955) undertook a study which compared educational administrators and aircraft commanders. A group of educational leaders and a group of B-29 and B-50 aircraft commanders were rated on the LBDQ by their subordinates and by themselves. This resulted in a comparison of "ideal" and "real" leader behavior. There was very little or no relation

between a leader's belief in how he should behave and his behavior as described by the group members. The hypothesis that the educational administrators, both in their leadership beliefs and leader behavior, would display more Consideration and less Initiation of Structure than the military leaders was supported. Since the administrators scored high on only the Consideration dimension, it was assumed that educational leaders were not as effective as leaders as they might have been.

Meanwhile, Fleishman, Harris and Burt (1955) were testing a modified version of the SBD (called the Foreman Behavior Description) in industrial settings. It appeared that more Consideration and less Initiating Structure on the part of foremen would be a good thing for everybody. The authors decided to find out if this assumption was valid. They used four objective criteria to measure foreman effectiveness. The objective criteria were absenteeism, accidents, grievances, and turnover. The resultant correlations indicated that high Consideration went with low absenteeism while high Initiation of Structure was accompanied by high absenteeism. There was also a tendency for high Initiation of Structure to go with increased grievances. Accidents and turnover had no appreciable relation to either dimension. As in an earlier study (Halpin, 1954), the subordinate workers preferred the foremen that were high on the Consideration dimension and low on Initiation of Structure. The most

efficient foremen, in the eyes of their superiors, were those who showed more Initiation and less Consideration.

Fleishman (1957) examined all the available evidence on the SBD, and summarized both the military and industrial findings. The Consideration and Initiating Structure dimensions were shown to be independent and reliable. Validity was assessed through correlations with independent leadership measures such as objective group indices (absenteeism and turnover), productivity ratings, peer ratings, and leaderless group situation tests. The scales were found differentially predictive of a number of these criteria. Correlations with other measures revealed that the scores achieved were independent of measures of general intelligence. Low correlations between the scales and other psychometric and background measures, and substantial correlation with certain leadership criteria suggested that these scales had strong potential as useful additions in the field of leadership research and assessment.

Fleishman and Peters (1962) expanded the use of the SBD in an industrial setting. The results were similar to previous military studies in that a manager's leadership attitudes toward the Consideration and Structure dimensions did not correspond with top management's effectiveness ratings. Higher level managers tended to feel that they should initiate less Structure with subordinates. Top management tended to identify the effectiveness of subordinate managers

with the effectiveness of their superiors. While subordinate managers could make their own bosses "look good," the reverse was also possible. Consideration and Structure were positively related only in the fact that those leaders who were most effective tended to emphasize both dimensions.

In a related study, Fleishman and Harris (1962) investigated the relationship between the leader behavior of industrial supervisors and the behavior of their group members. They found that low Consideration and high Structure went with high grievance and turnover rates. There appeared to be certain critical levels beyond which increased Consideration or decreased Structure had no effect on grievance or turnover rates. Foremen could compensate for high Structure by increased Consideration, but low Consideration foremen could not compensate by decreasing their Structuring behavior. A most significant finding was that leader behavior tended to be quite stable over a period of years. This stability removed some of the possible variability in the measures and enhanced the results relative to previous findings.

Hills (1963) determined that the Consideration and Initiating dimensions were not solely concerned with problems of internal leadership. His findings showed that consideration for group members could also be reflected in the manner with which a leader deals with outsiders, and that a leader could initiate structure upward as well as downward. Thus, the two dimensions were much more generalizable than

had been previously thought--having external as well as internal applicabilities.

Only one study (Mitchell, 1970) questioned the reliability and validity of the LBDQ and SBD. This study was unique in that it used a new statistical technique called the Multitrait-multimethod Analysis of Validity. The results indicated that the variables under examination (Initiation and Consideration) had relatively low reliabilities and an almost complete lack of relative construct validity. These findings did not cast any real doubt upon the numerous studies cited previously, especially in view of the fact that this research involved ratings of a very short interaction situation by ad hoc group members. The leaders of these groups were also arbitrarily assigned rather than elected or appointed by management. Deficiencies of this kind made the credibility of the entire study doubtful.

Thus, the lengthy and painstaking research on leadership seems to have produced a useful questionnaire which delineates two dimensions along which leader behavior varies.

Having established the measuring instrument used as psychometrically sound, another variable, peer assessment of leadership style, was introduced. Several previous studies (Bass, 1957; Fleishman, 1957a, 1957b; Spitzer & McNamara, 1964) had used peer ratings as criteria, but none utilized peer ratings to describe or rate leadership style.

Peer ratings, and the findings attached thereto, have primarily come from military settings. Once again, the military situation provided the opportunity for group processes and interaction to be investigated under relatively controlled conditions in a minimum of time.

Wherry and Fryer (1949) investigated the criticism that buddy ratings were not criteria of leadership but merely popularity contests. They found stability in peer ratings in that ratings made during the first month measured the same factors three months later. In addition, it was not until the fourth month that superiors' ratings reflected the leadership factors which fellow students identified in their first month ratings. Similar results were obtained by a number of other investigators where stability, validity, and reliability were concerned (Borgatta & Sperling, 1963; Fiske & Cox, 1960; Gordon & Medland, 1965; Hoffman & Rohrer, 1954; Hollander, 1957; Suci, Vallance & Glickman, 1954).

Hollander (1956) found that the validity of peer nominations was not adversely affected by considerations of friendship, while Lewin, Dubno, and Akula (1971) came to the conclusion that face-to-face interaction is not a critical variable in the peer rating process. Sells, Trites, and Parish (1957) found an interesting correlate of manifest anxiety and peer ratings where student pilots with high anxiety tended to receive lower buddy ratings from classmates and vice versa. Anxiety later correlated strongly

with class standing at graduation with high anxiety students graduating in the lower half of the class.

A few studies (Bartlett, 1959; de Leon, de Leon & Swihart, 1969; Werdelin, 1966) examined the relationship between peer ratings and self ratings and found the former to be a much more stable and reliable measure.

By far, the largest area of research on peer ratings was concerned with predictive attributes. Peer ratings have been successfully used to predict success in military flight training (Hollander, 1954b; Wahlberg, Boyles, & Boyd, 1971; Willingham, 1958), success as a supervisor or leader (Amir, Kovarsky & Sharan, 1970; Kraut, 1969; Reynolds, 1966; Roadman, 1964; Weitz, 1958), failure to complete military flight programs (Doll & Longo, 1962; Doll, 1963), success as a future naval officer (Hollander, 1965), success or failure as a member of the Peace Corps (Grande, 1966), first-year grade-point average in graduate school (Wiggins, Blackburn & Hackman, 1969), and success as a salesman (Waters & Waters, 1970).

In addition, peer ratings have been reviewed and found useful in an even wider variety of situations (Forehand & Guetzkow, 1961; Guion, 1965, pp. 472-473; Hollander, 1954a; Holmes, 1971). Peer ratings then are reliable, stable, and valid, and have had sufficient use to be considered sound for assessment purposes.

The relationship between style of leadership and group performance has received a very small portion of the

leadership research effort. As early as 1950, Katz, Maccoby, and Morse (1950) had found that the supervisors of high-producing sections seemed to think and act differently with respect to their supervisory functions than did supervisors of low producing sections. These results could have been the effect of productivity differences--that is, the section heads of low producing employees could have become "production centered" because their employees were doing poorly. Their attempts to supervise more closely might have been reinforced or even initiated by their superiors who had become concerned about low productivity. High producing sections would have received opposite treatment by both their superior and his superior. It seemed logical that differences in group motivation were related to differences in supervisory practice and philosophy. In a more recent study (Lowin & Craig, 1968), the level of group performance was manipulated and found to shape the leader's style of management. Performance changed supervision instead of vice versa.

Lawshe and Nagle (1953) established that the supervisor's behavior, as perceived by the employees, was closely related to the output of the work group. Similar findings were evident in a number of other studies as well (Day & Hamblin, 1964; Fleishman, et al., 1955; Greer, 1961; Wyndham & Cooke, 1964). All of this research was conducted in civilian industrial populations. None of it included military leaders, in particular the leaders of flying crews.

More recently, Cummins (1971) found support for the hypothesis that the effect of Initiating Structure on related productivity and quality of work-group performance, in an industrial manufacturing plant, was moderated by considerate leader behavior. The results supported the hypothesis when quality was the criterion measure, but not when productivity was the criterion measure. Thus, what research has been related to productivity is sketchy and inconclusive.

One purpose of the present study was to derive and use a better measure of productivity. Ghiselli (1956) indicated that any single dimension of job performance could not be taken as synonymous with job performance in its entirety. Performance on any job was said to be best described in terms of at least several independent dimensions. Most efforts to measure performance had been limited to the use of one measuring mode which was expected to measure total job performance.

Taylor, Parker, and Ford (1959) questioned the ability of the average supervisor to provide an appraisal of his subordinates adequate enough to be used as the true measure of performance. Supervisor ratings were seen as a measure of only one aspect of job performance. Many other aspects remained.

Kipnis (1960) advocated the development of other measures of job performance without the expectation that the measures need correlate with each other, or more importantly,

correlate with supervisor's ratings. Instead, she recommended that if technical knowledge was an important component of the criterion, job knowledge tests might be used; if willingness and diligence were important, supervisor ratings could be included; if human relations skills were to be measured, peer or subordinate ratings could be used. It seemed that the more job aspects measured, the better the productivity estimate, and in spite of Kipnis' warnings as to the difficulty of obtaining such data, the study described in this paper attempted to include as many job aspect measures as possible.

In the present study peers assessed each other's leader behavior via the SBD, and the two dimensional scores obtained therefrom were related to an improved productivity measure. It was assumed that a leader's style would manifest itself in his own productivity as well as in the performance of his subordinates. Therefore, each leader's own productivity was the criterion measure used.

Since peers had no reason to view each other as either superiors or subordinates did, it appeared probable that peer assessment of leader behavior would differ somewhat from these other ratings. Peers would not look primarily for high Consideration, as subordinates did, nor should they be concerned only about Structure, as superiors were. A peer would be in an ideal position to observe and rate his equals on whatever it was that the observed person possessed, or did not possess as a leader, without being coerced into looking for any single

dimension of style dictated by a position above or below that person. Hills (1963) demonstrated that a person's style of leadership was evident not only in relations with superiors and subordinates, but also in the manner in which a leader dealt with outsiders and non-affiliates of his organization.

With these findings in mind, three hypotheses were put forth to be tested. Hypothesis 1 proposed that those leaders rated by their peers as being high on Consideration would have significantly higher productivity scores than those rated as being low on Consideration. Hypothesis 2 proposed that those leaders rated by their peers as being high on Structure would have significantly higher productivity scores than those rated as being low on Structure. Hypothesis 3 proposed that those leaders rated by their peers as being high on both Consideration and Structure would have significantly higher productivity scores than any other group of leaders.

Method

Subjects

To obtain groups of men who interacted often and performed as leaders in the presence of each other on a daily basis, the U.S. Air Force training system, more specifically the pilot training program, was decided upon.

The Ss were 61 instructor pilots from a U.S. Air Force undergraduate pilot training base. They ranged in age from twenty-two to twenty-eight, and possessed similar educational

backgrounds, with all holding bachelor's degrees and three holding master's degrees.

Procedure

The typical pilot training base (wing) consisted of two training squadrons, each of which was made up of between six and eight training flights. Each flight consisted of between ten and twelve instructor pilots, each of whom had three to five student pilots assigned to him. Each flight also had a flight commander who was superior to the instructor pilots and their students and responsible for the total operation of the flight.

The training flights were very close groups, and members came to know one another very well in a short period of time. The daily interaction between instructor pilots was considerable, and each had many opportunities to observe the leader behavior of the others. Since each instructor and his students sat at tables in a large, open flight room, and since many syllabus missions required instructors to brief and fly with each other daily, each had multiple occasions to observe what kind of leaders the others were both on the ground and in flight.

The flight commander, also a qualified instructor pilot, was in an excellent position to observe the leadership behavior of all the instructors since he flew with and observed all of them routinely. All instructors flew two-engine jet trainers which carried one instructor pilot and one student

pilot per aircraft. The course of training was specifically outlined in a syllabus which included four major areas of concentration--contact (visual flight), instrument flying, formation (multiple aircraft flown close together), and navigation.

A slightly modified form of the SBD was used. The original SED consisted of 48 items 20 of which measured the Structure dimension and 28 of which measured the Consideration dimension. Since this questionnaire was designed for subordinate ratings of supervisor behavior, a number of items were worded inappropriately for peer ratings. Ten items were eliminated leaving 21 Consideration items and 17 Structure items for a total of 38 items (see appendix).

The use of shortened versions of the SBD had been demonstrated in a number of studies which sought to decrease the time involved in filling out the questionnaire. Stogdill and Shartle (1955) showed that reliabilities for descriptions by other persons ranged from .51 to .85 for the separate scales, and .89 for the total score. These reliabilities were as high as could be expected with each scale composed of only two to six items. Hemphill (1957) and Halpin (1957a) used questionnaires of 30 items in length (15 on each dimension) with highly satisfactory results and reliability. The elimination of the 10 inappropriate items then was not seen as an important factor in the study.

A total of 110 instructor pilots, representing 80% of the total wing, each filled out three modified SBD's after being read the following instructions.

Gentlemen:

In a couple of minutes I will give each of you three questionnaires to fill out. I want you to fill the questionnaires out on three instructor pilots in your flight. Please do not rate the flight commander or any of the staff personnel--use only your peers. Each of you observes the leadership behavior of your fellow flight members daily. I want you to select the three members of your flight that you feel most qualified to rate on their leadership behavior. This is not intended to be a popularity or personality contest, so please be as objective as you can in your selections. Select those people whose leader behavior you are most familiar with.

Do not fill out anything on the front page--when you begin to fill each questionnaire out just print the name of the person you are rating near the center fold. The names are necessary only to allow later correlation of the data. I guarantee that each of you will remain completely anonymous--name, squadron, and base will not be used in later analysis.

I want to caution you that there are no right or wrong answers. The items simply describe the leadership behavior of an individual; they do not judge whether his behavior is desirable or undesirable. Differences are expected, so please do not hesitate to mark an extreme answer if it is appropriate, or right down the middle if that is appropriate.

When you open the questionnaire you will notice that some of the items are crossed out. Ignore these items. Fill out only the unmarked items. They are simple multiple choice and require only a check mark for a response. Again, make sure each questionnaire has a name in it.

Are there any questions?

The purpose of having each instructor fill out questionnaires on three of his peers was to obtain some overlap in ratings. Though 85 instructors had questionnaires filled out on them, only 64 of these met the criterion of having three or more SBD's pertaining to them. Three of these

potential Ss were eliminated when they did not meet the criterion of having been in their present flight for a minimum of six months.

The 61 remaining S's SBDs were scored via the standard method of separately totalling the two dimensional scores. Each alternative on the questionnaire had five selections which were scored 0, 1, 2, 3, 4, or the reverse, whichever was appropriate. Each item and its appropriate dimension then had a mean of 2.00. Both the Consideration and Structure dimensions were dichotomized at the mean into high and low segments. A mean Consideration or Structure score of 2.00 or more was regarded as being high on the dimension. Mean scores below 2.00 were regarded as low. In this manner, Ss were assigned to one of four possible groups--high on both dimensions, low on both dimensions, high Consideration/low Structure, and low Consideration/high Structure. It was ascertained, before any questionnaires were filled out, that instructors rarely were aware of any peer evaluation or test scores. The Ss were unaware of the selected performance criteria. Flight commanders were aware of such scores.

The productivity measure used was a composite of five separate pieces of information considered relevant to an instructor pilot's performance. Each instructor was given an annual flying evaluation in each of four areas of emphasis which encompassed skills necessary for the training

program. These major areas were contact flying, instruments, formation, and navigation. On each of these evaluation flights, the instructor pilot flew with a member of the wing standardization board which was responsible for maintaining the professionalism and proficiency of all wing instructor pilots. Though it was possible for the same evaluator to fly more than one of these annual check flights with any individual instructor, it was most common that different evaluators were used. This procedure was followed to reduce the effects of personal biases or personality conflicts, and to maintain as much objectivity as possible.

These evaluation flights put the evaluatee in the role of an instructor pilot while the evaluator assumed the role of a student pilot. A typical mission profile was briefed and flown by the instructor while the evaluator graded both instructional ability and flying proficiency. Instructional ability was a reflection of how the instructor briefed and prepared the student for the mission, directed and planned maneuvers, analyzed poor performance, and debriefed after the mission. Flying proficiency scores were a reflection of actual aircraft handling, demonstration of maneuvers, and patterns and landings.

The instructional scores from the most recent evaluations, in the four areas of emphasis, were averaged for each instructor and this constituted the first aspect of productivity. The mean proficiency scores constituted the second.

Both the instructional and proficiency scores were in the form of points possible out of 100 carried to two decimal places. The instructional scores ranged from 79.50 to 90.25 with a mean of 85.23. The proficiency scores ranged from 79.50 to 91.00 with a mean of 84.84.

A third aspect of productivity was the number of errors given during the flight evaluations. An error was an objectively defined mistake or oversight on the part of the instructor which could have led to a compromise of flight safety, possible damage to equipment, or violation of flying procedures and regulations. Errors were assessed over and above instructional and proficiency scores and affected neither.

Total number of errors assessed per instructor during the most recent evaluations in the four emphasis areas and their assigned weights are presented in Table 1.

Table 1

Weights Assigned Relative to Total Errors Assessed	
Total Errors	Weight Value
0 - 2	0.0
3 - 5	-1.0
6 - 7	-2.0
8 - 9	-3.0
10 - 11	-4.0

A total of three or more errors on any single evaluation constituted failure of that mission. Since errors represented a negative aspect of performance, they had to be weighted and subtracted from the total productivity score.

The fourth aspect of productivity was supervisor ratings. Each flight commander was asked to force-rank the instructors in his flight in relation to his opinion of them as leaders. They were cautioned not to put too much emphasis on any one individual attribute such as skill as a pilot or instructor, but rather to assess each person more globally as a leader of men. The rankings and corresponding weights are presented in Table 2.

Table 2

Weights Assigned Relative to Flight Commander Ranking	
Rank Position	Weight Value
1	+4.00
2	+3.00
3	+2.00
4-6	+1.00
7-10	0.00

Though any number of weight values could have been assigned to the total error and rank position groups described, the values actually used were considered to be very conservative and fair. The weights increase more or less linearly

and were not so high that the composite productivity scores were greatly affected. This was desirable so that a single poor mission could not result in a heavy docking of performance points. Moreover, it was quite common for the average instructor to acquire an error on any given evaluation, and it would therefore have been unfair to subtract heavily for relatively few errors. Flight commander rankings were the only subjective measure of performance used and for that reason were weighted conservatively also. Thus, errors and rankings did adjust the performance composite score, but only in a reasonable and non-obscuring manner.

Another aspect of productivity was reflected in written job knowledge tests. Such tests were often given in conjunction with the flying evaluations. However, no-notice and higher headquarters tests were administered on other occasions. These tests measured knowledge of aircraft systems, emergency procedures, operating procedures, flying regulations, and a host of other topics. The four most recent written test scores were averaged for each instructor as a measure of the fifth aspect of performance. Written test scores ranged from 80.30 to 96.00 with a mean of 89.28.

The five aspect scores were then put together in the following manner. The mean instructional, proficiency, and written test scores were averaged. Error weights were then subtracted, and flight commander weights added to this mean to yield the overall productivity measure. Two calculation

examples are presented in Table 3 which demonstrate how the specific aspect inputs were combined for the highest and lowest overall productivity scores obtained in the study.

Table 3
Sample Calculations of Highest and Lowest
Overall Productivity Scores

Example Score	Productivity Aspect Inputs						Overall Productivity Score
	Mean of I Scores	Mean of P Scores	Mean of W Scores	Mean of I, P, & W Scores	Total Errors Weight	Rank Weight	
Highest	89.00	89.75	93.50	90.75	$\frac{2}{0.00}$	$\frac{1}{+4.00}$	94.75
Lowest	80.75	81.50	80.25	80.83	$\frac{11}{-4.00}$	$\frac{9}{0.00}$	76.83

I = Instructional Scores
P = Proficiency Scores
W = Written Test Scores

Results and Discussion

A summary of the data obtained is presented in Table 4.

Table 4
Number of Subjects and Mean Productivity
by Leadership Style

Structure	Consideration		Row Means
	Low	High	
High	$\bar{X}_1 = 84.2950$ $n = 12$	$\bar{X}_2 = 88.9152$ $n = 25$	87.4168
Low	$\bar{X}_3 = 82.8000$ $n = 8$	$\bar{X}_4 = 86.3325$ $n = 16$	85.1550
Column Means	83.6970	87.9073	86.5269

A 2 × 2 analysis of variance for unequal N_s , using the method of unweighted means (Winer, 1971, pp. 402-404), indicated differences for both main effects which could not be attributed to chance, and no significant interaction of the two main effects. A summary of the analysis of variance is presented in Table 5.

Table 5
Summary of Analysis of Variance

Source	df	MS	F
Structure (A)	1	53.4939	6.8285*
Consideration (B)	1	213.8334	27.2960**
A × B	1	3.8062	0.4859

* $p < .05$
** $p < .01$

These results supported Hypotheses 1 and 2 fully in that being high on either dimension went with being significantly more productive. Peers did not find it necessary to look for primarily one dimension as Halpin (1954) found to be true of subordinates and superiors. Being neither a subordinate nor a superior left a man's equals more likely to see what he had, not what they wished him to have. Whatever leadership skills were present, whether it was high Structure, high Consideration or both, were seen as assets. While some individuals were adept at human relations, others were more oriented toward planning, trying and structuring, but both

approaches were seen as being valuable aspects of leadership by themselves.

The preceding observations could have been interpreted as a manifestation of halo effect since peers did appear to have the same view of a ratee's abilities as the evaluators did. This could have been possible due to the personality of the ratee or the manner in which he approached problems; however, serious halo visibility was considered remote in this case. First, it must be remembered that the raters were not aware of their peer's productivity scores. Second, the wording and approach of the individual SBD items was such that halo effect was minimized (see Appendix).

If being high on either dimension was related to being significantly more productive than being low, then being high on both dimensions should have been significantly more productive than all the other groups combined. That was the contention of Hypothesis 3.

Since there was no significant interaction in the analysis of variance, a definitive statistical test of Hypothesis 3 was not considered appropriate. However, an inspection of Table 4 does lend support to the notion that those leaders rated by their peers as being high on both dimensions had higher productivity scores than any other group of leaders. Though significance cannot be inferred from this inspection of means, it is evident that Hypothesis 3 is supported at least superficially. The mean productivity scores do not

show evidence that would cause one to totally reject the hypothesis where subsequent research is concerned.

Support for the hypotheses proposed was the major finding of this study. However, there were several ancillary *erving as an aid* results which deserve mention. First, the SBD and the two dimensions that it measures appear to still be valid and useful in the field of leadership research. The current findings fully support previous results obtained using this questionnaire. Second, it seems that the use of multiple inputs into the criterion measure resulted in findings consistent with previous research, and in which one can place more confidence. The more objective multiple criterion measure, though difficult to gather, is more representative and comprehensive than previously used subjective measures. The multiple objective criterion approach to measures of productivity represents a definite step forward in group/leadership research.

Peer assessment of leader behavior is in line with previous findings which used subordinate and superior ratings. Peer ratings were different only in that peers were not under pressure to rate on the basis of only Consideration as subordinates had or only Structure as superiors had. Peers seemed to be more objective in assessing a person's style of leadership, and as Wherry and Fryer (1940) found, they could do it in a shorter period of time with more accuracy.

One drawback of research of this type is that it was after the fact. It would be interesting and highly useful to use peer assessment of leadership as a predictive device. A good example of this would be a longitudinal study in the pilot training situation. Student pilots could rate each other's style of leadership after several months in training. Those who returned to become instructor pilots later could then have productivity data gathered on them. Style of leadership would then be compared to productivity scores as in the present study. In this manner, a measure of predictive validity could be obtained. Proper refinement of this could result in a selection/rejection tool for use in the future.

Another question that was prompted by this study pertained to how generalizable the results would be to civilian situations. Although there was no basis to assert that civilian leaders would behave in the same ways, it could easily be assumed that human leaders in fairly structured situations would behave similarly as many structured situations exist in industry and education as well as in the military.

It should also be remembered that military populations are especially suited to this type of research in that large numbers of subjects are available in widespread locations, a wide variety of jobs are available for study, military groups are organized quickly under a variety of psychological and environmental situations, and most importantly, a fairly objective pool of performance/productivity data is kept. It

is unfortunate for this type of research that many civilian organizations do not keep the performance records that military sources do.

Since the present study assumed that a person's style of leadership would manifest itself in one's own productivity as well as in one's subordinates' performance, it appeared that one other logical research step is open to future study.

A situation is needed where subordinate performance can be measured and compared to the leader's performance to ascertain the correlation between them. In this way, the validity of the assumption made in this study could be tested. Certainly there must be some correlation between a leader's performance and the performance of his crew, but whether the relationship is significant or not would rest with the results of such a study.

Another area of question, to be resolved by subsequent research, rests in the possible relationship of subordinate, peer, and superior ratings. For example, a fairly large study of bomber crews using subordinate, peer, and superior ratings would be able to ascertain if peers see the same people being categorized by a style of leadership that subordinates or superiors do. The present study does give prima-facie evidence that this is the case, but not empirical data to that effect. Coupled with an improved multifaceted productivity measure, the results of such a study could be very revealing.

The ultimate purpose of the type of research described in this study is to obtain a means of predicting whether or not a given person will be effective as a leader, and what constitutes "effectiveness." Before prediction can be attempted, valid and reliable measurement of leadership style is necessary. Who is best to rate a person's style of leadership also must be known. This is the position of research in leadership at the present time, and much more is needed before we could assume to know enough to select or reject on the basis of a predictive device. It is hoped that the results of this study represent a positive step in resolving these issues.

APPENDIX
SUPERVISORY BEHAVIOR DESCRIPTION

SUPERVISORY BEHAVIOR DESCRIPTION

by

Edwin A. Fleishman, Ph.D.
 American Institutes for Research
 Washington, D.C.

Name _____ Date _____
 _____ (Middle) _____
 _____ (First) _____ Position _____
 _____ (Last) _____ Company _____

Raw Score	Percentile	Other
C		
S		
Description of Norm Group		

INSTRUCTIONS:

You have observed your own supervisor and probably you know pretty well how he operates. In this questionnaire, you are simply to *describe* some of the things your own supervisor does with your group.

For each item, choose the alternative which best describes how often your supervisor does what that item says. Remember...there are no right or wrong answers to these questions. The items simply *describe* the behavior of the supervisor over you; they do not judge whether his behavior is desirable or undesirable. Everyone's supervisor is different and so is every work group, so we expect differences in what different supervisors do.

Answer the items by marking an "X" in the box (a, b, c, d or e) next to each item to indicate your choice.

Copyright 1970, Edwin A. Fleishman, Ph.D.

Printed in U.S.A.

All rights reserved.

USAF SCN 72-59

1. HE IS EASY TO UNDERSTAND. a b c d e
a. always b. often c. occasionally d. seldom e. never
2. HE ENCOURAGES OVERTIME WORK. a b c d e
a. a great deal b. fairly much c. to some degree d. comparatively little e. not at all
3. HE TRIES OUT HIS NEW IDEAS. a b c d e
a. often b. fairly much c. occasionally d. once in a while e. very seldom
4. HE BACKS UP WHAT PEOPLE IN HIS WORK GROUP DO. a b c d e
a. always b. often c. occasionally d. seldom e. never
5. HE CRITICIZES POOR WORK. a b c d e
a. always b. often c. occasionally d. seldom e. never
- ~~6. HE DEMANDS MORE THAN WE CAN DO. a b c d e
a. often b. fairly often c. occasionally d. once in a while e. very seldom~~
7. HE REFUSES TO GIVE IN WHEN PEOPLE IN THE WORK GROUP DISAGREE WITH HIM. a b c d e
a. always b. often c. occasionally d. seldom e. never
- ~~8. HE EXPRESSES APPRECIATION WHEN ONE OF US DOES A GOOD JOB. a b c d e
a. always b. often c. occasionally d. seldom e. never~~
9. HE INSISTS THAT PEOPLE UNDER HIM FOLLOW STANDARD WAYS OF DOING THINGS IN EVERY DETAIL. a b c d e
a. always b. often c. occasionally d. seldom e. never
10. HE HELPS PEOPLE IN THE WORK GROUP WITH THEIR PERSONAL PROBLEMS. a b c d e
a. often b. fairly often c. occasionally d. once in a while e. very seldom
11. HE IS SLOW TO ACCEPT NEW IDEAS. a b c d e
a. always b. often c. occasionally d. seldom e. never
12. HE IS FRIENDLY AND CAN BE EASILY APPROACHED. a b c d e
a. always b. often c. occasionally d. seldom e. never
- ~~13. HE GETS THE APPROVAL OF THE WORK GROUP ON IMPORTANT MATTERS BEFORE GOING AHEAD. a b c d e
a. always b. often c. occasionally d. seldom e. never~~
14. HE RESISTS CHANGES IN WAYS OF DOING THINGS. a b c d e
a. a great deal b. fairly much c. to some degree d. comparatively little e. not at all
- ~~15. HE ASSIGNS PEOPLE UNDER HIM TO PARTICULAR TASKS. a b c d e
a. always b. often c. occasionally d. seldom e. never~~
16. HE STRESSES BEING AHEAD OF COMPETING WORK GROUPS. a b c d e
a. a great deal b. fairly much c. to some degree d. comparatively little e. not at all
17. HE CRITICIZES A SPECIFIC ACT RATHER THAN A PARTICULAR INDIVIDUAL. a b c d e
a. always b. often c. occasionally d. seldom e. never

18. HE LETS OTHERS DO THEIR WORK THE WAY THEY THINK BEST. a b c d e
a. always b. often c. occasionally d. seldom e. never
19. HE DOES PERSONAL FAVORS FOR THE PEOPLE UNDER HIM. a b c d e
a. often b. fairly often c. occasionally d. once in a while e. very seldom
20. HE EMPHASIZES MEETING OF DEADLINES. a b c d e
a. a great deal b. fairly much c. to some degree d. comparatively little e. not at all
21. HE SEES THAT A WORKER IS REWARDED FOR A JOB WELL DONE. a b c d e
a. always b. often c. occasionally d. seldom e. never
22. HE TREATS PEOPLE UNDER HIM WITHOUT CONSIDERING THEIR FEELINGS. a b c d e
a. always b. often c. occasionally d. once in a while e. very seldom
- ~~23. HE INSISTS THAT HE BE INFORMED ON DECISIONS MADE BY THE PEOPLE UNDER HIM. a b c d e
a. always b. often c. occasionally d. seldom e. never~~
24. HE OFFERS NEW APPROACHES TO PROBLEMS. a b c d e
a. often b. fairly often c. occasionally d. once in a while e. very seldom
- ~~25. HE TREATS ALL WORKERS UNDER HIM AS HIS EQUALS. a b c d e
a. always b. often c. occasionally d. seldom e. never~~
26. HE IS WILLING TO MAKE CHANGES. a b c d e
a. always b. often c. occasionally d. seldom e. never
27. HE ASKS SLOWER PEOPLE TO GET MORE DONE. a b c d e
a. often b. fairly often c. occasionally d. once in a while e. very seldom
28. HE CRITICIZES PEOPLE UNDER HIM IN FRONT OF OTHERS. a b c d e
a. often b. fairly often c. occasionally d. once in a while e. very seldom
29. HE STRESSES THE IMPORTANCE OF HIGH MORALE AMONG THOSE UNDER HIM. a b c d e
a. a great deal b. fairly much c. to some degree d. comparatively little e. not at all
30. HE TALKS ABOUT HOW MUCH SHOULD BE DONE. a b c d e
a. a great deal b. fairly much c. to some degree d. comparatively little e. not at all
31. HE "RIDES" THE PERSON WHO MAKES A MISTAKE. a b c d e
a. often b. fairly often c. occasionally d. once in a while e. very seldom
- ~~32. HE WAITS FOR PEOPLE UNDER HIM TO PUSH NEW IDEAS BEFORE HE DOES. a b c d e
a. always b. often c. occasionally d. seldom e. never~~
33. HE RULES WITH AN IRON HAND. a b c d e
a. always b. often c. occasionally d. seldom e. never
34. HE TRIES TO KEEP THE PEOPLE UNDER HIM IN GOOD STANDING WITH THOSE IN HIGHER AUTHORITY. a b c d e
a. always b. often c. occasionally d. seldom e. never

35. HE REJECTS SUGGESTIONS FOR CHANGES. a b c d e
a. always b. often c. occasionally d. seldom e. never
- ~~36. HE CHANGES THE DUTIES OF PEOPLE UNDER HIM WITHOUT FIRST TALKING IT OVER WITH THEM. a b c d e
a. often b. fairly often c. occasionally d. once in a while e. very seldom~~
37. HE DECIDES IN DETAIL WHAT SHALL BE DONE AND HOW IT SHALL BE DONE. a b c d e
a. always b. often c. occasionally d. seldom e. never
38. HE SEES TO IT THAT PEOPLE UNDER HIM ARE WORKING UP TO THEIR LIMITS. a b c d e
a. always b. often c. occasionally d. seldom e. never
39. HE STANDS UP FOR PEOPLE UNDER HIM EVEN THOUGH IT MAKES HIM UNPOPULAR. a b c d e
a. always b. often c. occasionally d. seldom e. never
40. HE MAKES THOSE UNDER HIM FEEL AT EASE WHEN TALKING WITH HIM. a b c d e
a. always b. often c. occasionally d. seldom e. never
- ~~41. HE PUTS SUGGESTIONS THAT ARE MADE BY THE PEOPLE UNDER HIM INTO OPERATION. a b c d e
a. always b. often c. occasionally d. seldom e. never~~
42. HE REFUSES TO EXPLAIN HIS ACTIONS. a b c d e
a. often b. fairly often c. occasionally d. once in a while e. very seldom
43. HE EMPHASIZES THE QUANTITY OF WORK. a b c d e
a. a great deal b. fairly much c. to some degree d. comparatively little e. not at all
44. HE ASKS FOR SACRIFICES FROM HIS PEOPLE FOR THE GOOD OF THE ENTIRE DEPARTMENT. a b c d e
a. often b. fairly often c. occasionally d. once in a while e. very seldom
- ~~45. HE ACTS WITHOUT CONSULTING THE PEOPLE UNDER HIM FIRST. a b c d e
a. often b. fairly often c. occasionally d. once in a while e. very seldom~~
46. HE "NEEDLES" PEOPLE UNDER HIM FOR GREATER EFFORT. a b c d e
a. a great deal b. fairly much c. to some degree d. comparatively little e. not at all
47. HE INSISTS THAT EVERYTHING BE DONE HIS WAY. a b c d e
a. always b. often c. occasionally d. seldom e. never
48. HE ENCOURAGES SLOW-WORKING PEOPLE TO GREATER EFFORT. a b c d e
a. often b. fairly often c. occasionally d. once in a while e. very seldom

REFERENCES

- ✕ Adorno, T., Frenkel-Brunswik, E., Levinson, D., & Sanford, R. N. The authoritarian personality. New York: Harper & Brothers, 1950.
- Amir, Y., Kovarsky, Y. & Sharan, S. Peer nominations as a predictor of multistage promotions in a ramified organization. Journal of Applied Psychology, 1970, 54, 462-469.
- Bartlett, C. J. The relationship between self-ratings and peer ratings on a leadership behavior scale. Personnel Psychology, 1959, 12, 237-246.
- ✕ Bass, B. Leadership opinions and related characteristics of salesmen and sales managers. In R. M. Stogdill & A. E. Coons (Eds.), Leader behavior: its description and measurement. Columbus, Ohio: Bureau of Business Research, 1957.
- Borgatta, E. F. & Sperling, B. C. A note of the stability of peer judgments in independent situations. Journal of Psychological Studies, 1963, 14, 45-47.
- Cummins, R. C. Relationship of initiating structure and job performance as moderated by consideration. Journal of Applied Psychology, 1971, 55, 489-490. ✓
- Day, R. C. & Hamblin, R. L. Some effects of close and punitive styles of supervision. The American Journal of Sociology, 1964, 69, 499-510.

- de Leon, P. H., de Leon, J. L. & Swihart, P. J. Relation of accuracy of self-perception and peer ratings. Perceptual and Motor Skills, 1969, 29, 966.
- Doll, R. E. & Longo, A. A. Improving the predictive effectiveness of peer ratings. Personnel Psychology, 1962, 15, 215-220.
- Doll, R. E. Peer rating validity as a function of rater intelligence and rating score received. U. S. N. School of Aviation Medical Research Report, 1963, 24, 5 p. (abstract).
- Fiske, D. W. & Cox, J. A., Jr. The consistency of ratings by peers. Journal of Applied Psychology, 1960, 44, 11-17.
- * Fleishman, E. A. The description of supervisory behavior. Journal of Applied Psychology, 1953, 37, 1-6. (a)
- * Fleishman, E. A. The measurement of leadership attitudes in industry. Journal of Applied Psychology, 1953, 37, 153-158. (b)
- * Fleishman, E. A., Harris, E. F. & Burtt, H. E. Leadership and supervision in industry. Columbus, Ohio: Bureau of Education Research, Ohio State University, 1955.
- * Fleishman, E. A. A leader behavior description for industry. In R. M. Stogdill & A. E. Coons (Eds.), Leader behavior: its description and measurement, Columbus, Ohio: Bureau of Business Research, Ohio State University, 1957. (a)

- Fleishman, E. A. The Leadership Opinion Questionnaire. In R. M. Stogdill & A. E. Coons (Eds.), Leader behavior: its description and measurement. Columbus, Ohio: Bureau of Business Research, Ohio State University, 1957. (b)
- Fleishman, E. A. & Harris, E. F. Patterns of leadership behavior related to employee grievances and turnover. Personnel Psychology. 1962, 15, 43-56.
- Fleishman, E. A. & Peters, D. R. Interpersonal values, leadership attitudes, and managerial success. Personnel Psychology, 1962, 15, 127-143.
- × Fleishman, E. A. Supervisory Behavior Description. Washington, D. C.: American Institutes for Research, 1970.
- Forehand, G. A. & Guetzkow, H. The Administrative Judgment Test as related to descriptions of executive judgment behaviors. Journal of Applied Psychology, 1961, 45, 257-261.
- Ghiselli, E. E. Dimensional problems of criteria. Journal of Applied Psychology, 1956, 40, 1-4.
- Gordon, L. V. & Medland, F. F. The cross-group stability of peer ratings of leadership potential. Personnel Psychology, 1965, 18, 173-177.
- Grande, P. P. The use of self and peer ratings in a Peace Corps training program. Vocational Guidance Quarterly, 1966, 14, 244-246.

- Greer, F. L. Leader indulgence and group performance. Psychological Monographs, 1961, 75, (35 Whole No. 516).
- Guion, R. M. Personnel Testing. New York: McGraw-Hill, 1965.
- Halpin, A. W. The leadership behavior and combat performance of airplane commanders. Journal of Abnormal and Social Psychology, 1954, 49, 19-22.
- Halpin, A. W. The leader behavior and leadership ideology of educational administrators and aircraft commanders. Harvard Educational Review, 1955, 25, 18-32.
- Halpin, A. W. The leader behavior and effectiveness of aircraft commanders. In R. M. Stogdill & A. E. Coons (Eds.), Leader behavior: its description and measurement. Columbus, Ohio: Bureau of Research, Ohio State University, 1957.
- Halpin, A. W. & Winer, B. J. A factorial study of the Leader Behavior Descriptions. In R. M. Stogdill & A. E. Coons (Eds.), Leader behavior: its description and measurement. Columbus, Ohio State University, 1957.
- * Hemphill, J. K. & Coons, A. E. Development of the Leader Behavior Description Questionnaire. In R. M. Stogdill & A. E. Coons (Eds.), Leader behavior: its description and measurement. Columbus, Ohio: Bureau of Business Research, Ohio State University, 1957.
- Hills, R. J. The representative function: a neglected dimension of leadership behavior. Administrative Science Quarterly, 1963, 8, 83-101.

- Hoffman, E. L. & Rohrer, J. H. An objective peer evaluation scale: construction and validity. Educational and Psychological Measurement, 1954, 14, 332-341.
- Hollander, E. P. Authoritarianism and leadership choice in a military setting. American Psychologist, 1953, 8, 368 (abstract).
- Hollander, E. P. Buddy ratings: military research and industrial implications. Personnel Psychology, 1954, 7, 385-393. (a)
- Hollander, E. P. Peer nominations on leadership as a predictor of the pass-fail criterion in naval air training. Journal of Applied Psychology, 1954, 38, 150-153. (b)
- Hollander, E. P. The friendship factor in peer nominations. Personnel Psychology, 1956, 9, 435-447.
- Hollander, E. P. The reliability of peer nominations under various conditions of administration. Journal of Applied Psychology, 1957, 41, 85-90.
- Hollander, E. P. Validity of peer nominations in predicting a distant performance criterion. Journal of Applied Psychology, 1965, 49, 434-438.
- Holmes, D. S. Conscious self-appraisal of achievement motivation: the self-peer rank method revisited. Journal of Consulting & Clinical Psychology, 1971, 36, 23-26.
- Katz, D., Maccoby, N. & Morse, N. C. Productivity, supervision and morale in an office situation--Part I. Ann Arbor, Michigan: Institute for Social Research, University of Michigan, 1950.

- Kipnis, D. Some determinants of supervisory esteem. Personnel Psychology, 1960, 13, 377-392.
- * Korman, A. K. Consideration, initiating structure, and organizational criteria--a review. Personnel Psychology, 1966, 19, 349-361.
- Kraut, A. I. Intellectual ability and promotional success among high level managers. Personnel Psychology, 1969, 22, 281-290.
- Lawshe, C. H. & Nagle, B. F. Productivity and attitude toward supervisor. Journal of Applied Psychology, 1953, 37, 159-162.
- Lewin, A. Y., Dubno, P. & Akula, W. G. Face-to-face interaction in the peer-nomination process. Journal of Applied Psychology, 1971, 55, 495-497.
- * Lowin, A. & Craig, J. R. The influence of level of performance on managerial style: an experimental object lesson in the ambiguity of correlational data. Organizational Behavior and Human Performance, 1968, 3, 440-458.
- Mitchell, T. R. The construct validity of three dimensions of leadership research. The Journal of Social Psychology, 1970, 80, 89-94.
- Parker, T. C. Relationships among measures of supervisory behavior, group behavior, and situational characteristics. Personnel Psychology, 1963, 16, 319-334.
- Reynolds, H. H. Efficacy of sociometric ratings in predicting leadership success. Psychological Reports, 1966, 19, 35-40.

- Roadman, H. E. An industrial use of peer ratings. Journal of Applied Psychology, 1964, 48, 211-214.
- X Rohde, K. J. The relation of authoritarianism of the aircrew member to his acceptance by the airplane commander. American Psychologist, 1951, 6, 323.
- Sells, S. B., Trites, D. K. & Parish, H. S. Correlates of manifest anxiety in beginning pilot trainees. Journal of Aviation Medicine, 1957, 28, 583-588.
- Spitzer, M. E. & McNamara, W. J. A managerial selection study. Personnel Psychology, 1964, 17, 19-40.
- Stogdill, R. M. & Coons, A. E. Leader behavior: its description and measurement. Columbus, Ohio: Bureau of Business Research, Ohio State University, 1957. ✓
- Suci, G. J., Vallance, T. R. & Glickman, A. S. An analysis of peer ratings I. U.S. Bureau of Naval Personnel Technical Bulletin, 1954, 54-59, iv, 13 p.
- Taylor, E. K., Parker, J. W., and Ford, G. L. Rating scale content IV. Predictability of structured and unstructured scales. Personnel Psychology, 1959, 12, 247-266.
- Wahlberg, J. L., Boyles, W. R. & Boyd, H. A. Peer ratings as predictors of success in military aviation. HumRRO Professional Paper, 1971 (Mar), No. 1-71, 16 p.
- Waters, L. K., & Waters, C. W. Peer nominations as predictors of short-term sales performance. Journal of Applied Psychology, 1970, 54, 42-44.

- X Weitz, J. Selecting supervisors with peer ratings. Personnel Psychology, 1958, 11, 25-35.
- Werdelin, I. Teacher ratings, peer ratings, and self ratings of behavior in school. Educational and Psychological Interactions, 1966, 11, 1-19.
- Wherry, R. J. & Fryer, D. H. Buddy ratings: popularity contest or leadership criterion? Personnel Psychology, 1949, 2, 147-159.
- Wiggins, N., Blackburn, M. & Hackman, J. R. Prediction of first-year graduate success in psychology: peer ratings. Journal of Educational Research, 1969, 63, 81-85.
- Willingham, W. W. A note on peer nominations as a predictor of success in naval flight training. U.S.N. School of Aviation Medicine Project Report, 1958, 14, 5 p. (abstract)
- Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1971.
- Wyndham, C. H. & Cooke, H. M. The influence of the quality of supervision on the production of men engaged on moderately hard physical work. Ergonomics, 1964, 7, 139-149.