# RESEARCH NOTES

UNANNOUNCED

Number 52-63

September 1949



TAG-PRS-RESEARCH NOTE-52-63

12 15

## AN EVALUATION OF JOB SITUATION TESTS IN THE MEASUREMENT OF LEADERSHIP

PJ-3404-16



PERSONNEL RESEARCH SECTION
PR AND P BR, TAGO

Research Note 52-63

# AN EVALUATION OF JOB SITUATION TESTS IN THE MEASUREMENT OF LEADERSHIP[1]

In April 1947, research psychologists in The Adjutant General's Office of the Department of the Army, undertook a long-range research program involving the development of instruments and procedures for the selection of potential leaders, and the evaluation of the curriculum and training procedures used in developing leadership ability. Leaders Courses had been established at 4 newly instituted Potential Leaders Schools.

At the present time, there are 6 Potential Leaders Schools, one at each training division. The stated objective of the Leaders Course is to select potential leaders early in their military careers, and to develop their leadership potential, by instruction, and by supervised guidance in the performance of duties as acting non-commissioned officers and instructors. The course is conducted in two phases, the first primarily academic, and lasting 3 weeks. In the second phase (of three weeks duration) the student is assigned to a training company as acting non-commissioned officer and is provided opportunity to apply the principles of leadership behavior learned in Phase I of the Course.

Enlisted men, admitted to the Leaders Course are largely men with no prior service, who are judged to have the capacity for leadership, on the basis of AGCT score, physical fitness, previous civilian record of leadership activities, entries on a Biographical Information Blank and on an Evaluation Report.

One of the first problems, undertaken by Department of the Army research psychologists, in close cooperation with the staffs at Potential Leaders' Schools in the field, was to construct instruments, and devise procedures, for evaluating leadership performance during training. The Leaders Reaction Test was one of the instruments developed and put into operational use.

The Leaders Reaction Test is a performance test, involving 20 specified leadership situations, administered in the field, under field conditions. The situations are of the general type included in leadership performance tests used by the British, and by the Office of Strategic Services in this country.

The Leaders Reaction Test was conceived and constructed largely through the efforts of Dr. J. B. Maller, then a member of the staff of PRS. Miss Sally Greenberg presently of PRS, planned the analysis of the results under the direction of Dr. Hubert Brogden.

Each situation in the LRT is participated in by a group of 4 men, and provides a basis for observing and rating performance. Ratings are recorded on an Observation Record, specially constructed for the test.

---

1/ Paper delivered by Mary Morton at APA, Denver, Colorado, 1949.

The LRT is in three parts. Parts I and III consist of two situations each. A leader is not designated. The task in each situation is presented, and the 4 men work out and execute the solution. The tasks allow for the emergence of spontaneous leadership behavior. For each situation, each man's performance is evaluated in terms of motivation, initiative, cooperation, endurance and overall performance.

Part I - Small Job Management; provides two situations, involving some definite task, such as building a bridge, or pitching a tent.

Part III - Emergency Management; provides two stress situations of an emergency nature, such as caring for casualties under fire.

In Part II - Designated Leader Situations; 16 situations, resembling battle conditions, are provided. In each of the 16 situations, one of the 4 men in the group is designated as leader. The situations were designed for the display of four types of behavior: persistence in the face of obstacles; maintaining security and avoiding trouble when necessary to accomplish a mission; control and cohesion; and reorganization and improvisation.

For each of the 16 designated leader situations in Part II, seven observations of the behavior of the leader are recorded: reaction time, comprehension, effectiveness, cohesion (control over men), cooperation elicited, consideration for the men, and an over-all rating. In addition to these 7 ratings for the leader, a rating is given the performance of each of the 3 followers in the group (+ if adequate, 0 if no impression, - if inadequate).

No man is designated as leader in all 16 situations. Each man acts as designated leader in 4 of the 16 situations and as follower in the remaining 12. Rotation of leader roles throughout the 16 situations is such that any possible effects of practice and fatigue are minimized.

The analysis of primary interest to the LRT is concerned with the reliability of LRT scores given 400 Leaders Course graduates now assigned to duties involving leadership responsibilities.

The 400 men were given ratings on the LRT by 2 observers. The rating given by an observer can be considered as a score. Thus, the correlation between the ratings of the 2 observers can be interpreted as the reliability of the method of scoring the LRT.

For the total test the estimated correlation between ratings of the observers (or the reliability of scoring) was found to be .75. Although for the purposes of this analysis, the 400 men were rated by 2 observers, in practice there is often but one observer. If in all actual field situations there were 2 observers, the reliability of scoring is estimated to be .86 (using Spearman-Brown correction).

Total test scores ranged from 125 to 300, with the mean score at 219 and a sigma of 26.09. For leaderless situations, reliabilities range from .62 to .71. Scores range from 0 to 25, with the means from 18.72 to 19.81 and sigmas

from 2.92 to 4.09. For the designated leader situations, reliabilities ranged from .39 to .78, with 5 in the 70's, 5 in the 60's, and 4 in the 50's. Scores range from 0 to 35, with means from 23.05 to 25.48, and sigmas from 5.0 to 6.9

Variation in the means on the 16 leader situations is small, only 2.4 score points between the lowest and the highest. This fact is evidence of the comparability of the 16 situations, a comparability which is necessary since every designated leader does not have a chance at all 16 situations, but at only 4 of the 16. Each man is made leader of a group in a specified order. For the 16 designated leader situations (situations 3-18, presented in sequence), the order in which men A, B, C, and D are leaders is as follows: Man A is leader for situations 3, 4, 17 and 18; man B for 5, 6, 15 and 16; C for 7, 8, 13 and 14; and D for situations 9, 10, 11 and 12. A comparison of the means of scores achieved by all men, assigned to these 4 groupings of situations, shows that the groupings are comparable. Mean scores for the groupings of situations are 24.66, 24.30, 24.38, and 24.36, respectively.

Examination of the scatterplots shows a marked piling up of scores at score points 20 (for the leaderless) and 28 (for the leader situations). These are scores which could be obtained if observers made a single + mark (denoting adequate performance) on every trait for each situation. This piling up at the single + mark point suggests that raters showed an unduly strong tendency to be noncommittal in their ratings. The single + point on the scale was undoubtedly interpreted as a neutral position.

Reliability of Observations of Followers

Men acting as followers in leader situations, were rated on "overall performance." For the 16 leader situations, reliability coefficients ranged from .20 to .48 (with 7 in the 20's, 4 in the 30's, and 5 in the 40's). This range of .26 to .48 is in distinct contrast to the .39 to .78 range for leadership scores, suggesting the possibility that observers were less interested in follower behavior. The reliabilities for follower scores is low enough to throw some doubt on their usefulness.

Inter-reliability Between Observations of Traits Within Situations

Although the analysis is still in process, I was able to obtain some data on the reliability of traits, before leaving Washington. Estimated average intercorrelations of scores on traits for each of the 20 situations were sufficiently low to suggest that observers appear to be able to distinguish and to evaluate the different traits being observed within a given situation. Estimated average intercorrelations between traits range from 39 to 59 (with Md coefficient at 50) for leader situations. For the four leaderless situations, intercorrelations were .40, .76, .18, and .52

Considering the reliability of the individual traits, "reaction time" shows consistently high reliability throughout the 16 leader situations. Many coefficients for "consideration" were quite low.

From the analysis presented, it can be concluded that the method of obtaining observations used in the Leaders Reaction Test gives reasonably satisfactory reliability. A lack of spread, due to a piling up of observations at a neutral point on the scale, suggests an undesirable tendency on the part of observers toward noncommittal evaluations. The behavior of followers tends to be less reliably evaluated than that of leaders.

Incidental to the evaluation of other measures of achievement in Leaders Courses the relationship of LRT to other measures of achievement was studied.

For 4 populations, involving 2000 Potential Leaders School graduates, correlations were computed between LRT scores and 13 other variables employed in the appraisal of success at the school. Other variables included ratings of faculty, associates, academic grades showing progress, and ratings of performances as acting NCO's in the latter part of training. The resulting correlations are low positive, none exceeding .21, and suggest that the variables operating in the LRT are relatively independent of the other 13 measures of achievement at the school. No consistently high correlation was obtained for any measure. These coefficients were obtained incidental to analyses made for other purposes and cannot be considered validity coefficients since the 13 other variables do not include all of the factors which the LRT was designed to measure.

PERSONNEL: Mary A. Morton, J. B. Maller

PREPARATION OF RESEARCH NOTE: September 1949