CONSTRUCT VALIDITY OF LEADER EFFECTIVENESS CRITERIA

Ronald G. Downey, Paul J. Duffy, and Samuel Shiflett

PERSONNEL AND MANPOWER TECHNICAL AREA

U. S. Army
Research Institute for the Behavioral and Social Sciences
June 1979

79 10 22 072
Approved for public release; distribution unlimited.
U. S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES
A Field Operating Agency under the Jurisdiction of the
Deputy Chief of Staff for Personnel

JOSEPH ZEIDNER
Technical Director

WILLIAM L. HAUSER
Colonel, U. S. Army
Commander

NOTICES

DISTRIBUTION: Primary distribution of this report has been made by ARI. Please address correspondence concerning distribution of reports to: U. S. Army Research Institute for the Behavioral and Social Sciences, ATTN: PERI-P, 5001 Eisenhower Avenue, Alexandria, Virginia 22333.

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U. S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.
Leadership effectiveness analysis of leadership criteria

Previous research on leadership effectiveness has often used criteria of one type and from one source. In this research, data on leadership effectiveness were collected with a variety of measures from different sources (leaders, peers, subordinates). Findings were analyzed for statistical convergence and by a multitrait-multimethod intercorrelation matrix. A satisfying degree of convergence but relatively little divergence appeared, and 47% of the variance was not accounted for by either one. Apparently, (Continued)
many measurement techniques or sources of data are essentially interchangeable. One cannot safely interpret results across different studies, however, because of the large amount of variance that remains unaccounted for by either convergent or divergent validity. The report is written for psychologists.
CONSTRUCT VALIDITY OF LEADER EFFECTIVENESS CRITERIA

Ronald G. Downey, Paul J. Duffy, and Samuel Shiflett

Submitted by:
Ralph R. Canter, Chief
PERSONNEL AND MANPOWER TECHNICAL AREA

Approved By:
E. Ralph Dusek
PERSONNEL AND TRAINING RESEARCH LABORATORY
Joseph Zeidner
TECHNICAL DIRECTOR

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES
5001 Eisenhower Avenue, Alexandria, Virginia 22333
Office, Deputy Chief of Staff for Personnel
Department of the Army

June 1979

Army Project Number
20162717A766

Leader Effectiveness Criteria

Approved for public release; distribution unlimited.
ARI Research Reports and Technical Papers are intended for sponsors of R&D tasks and other research and military agencies. Any findings ready for implementation at the time of publication are presented in the latter part of the Brief. Upon completion of a major phase of the task, formal recommendations for official action normally are conveyed to appropriate military agencies by briefing or Disposition Form.
This research, carried out in the Personnel and Manpower Technical Area of the Army Research Institute for the Behavioral and Social Sciences (ARI), evaluated several modes of estimating leader effectiveness for comparability of the data produced by each.

The research was an in-house effort, responsive to Army Project 2Q162717A766 and to special requirements of the Office of the Deputy Chief of Staff for Personnel. At the time the research was conducted, both Samuel Shiflett and Ronald G. Downey were part of the ARI research staff. Dr. Shiflett is now at New York University, New York City, and Dr. Downey at Kansas State University, Manhattan, Kans. Robert Sulzen helped arrange and coordinate the data collection and provided insights into the group processes during the exercise. Frank E. Saal gave helpful comments on the paper.

JOSEPH ZEIDNER
Technical Director
CONSTRUCT VALIDITY OF LEADER EFFECTIVENESS CRITERIA

BRIEF

Requirement:

To investigate whether different measures of leadership effectiveness from different sources provide comparable data.

Procedure:

Data on leader effectiveness were gathered during a 2-week field exercise from three different sources (leaders, peers, and subordinates), using several different techniques (questionnaires on leader performance, leader self-ratings, and peer evaluation of leader effectiveness). These data were analyzed in a multitrait-multimethod fashion.

Findings:

The rating situation and raters' perceptions interact with leaders' behavior and evaluations. Most types of measuring techniques seemed to be evaluating the same basic variables, with the exception of peer evaluation of leadership potential. Exact amounts of interchangeability could not be determined.

Utilization of Findings:

The data suggested that different methods of measuring leadership may be subject to different kinds of bias, and that using more than one method and source in the assessment may be essential in measuring all aspects of leader performance.
## CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>METHOD</td>
<td>2</td>
</tr>
<tr>
<td>Subjects</td>
<td>2</td>
</tr>
<tr>
<td>Procedure</td>
<td>2</td>
</tr>
<tr>
<td>RESULTS</td>
<td>2</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>6</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>9</td>
</tr>
<tr>
<td>DISTRIBUTION</td>
<td>11</td>
</tr>
</tbody>
</table>

## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Item content of leader effectiveness, group performance, and group relations measures</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>Intercorrelations of different methods of measuring leader effectiveness</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>Multitrait-multimethod matrix of measures of leader effectiveness, group performance, and group relations</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>Analysis of variance of correlations</td>
<td>7</td>
</tr>
</tbody>
</table>
CONSTRUCT VALIDITY OF LEADER EFFECTIVENESS CRITERIA

INTRODUCTION

Research over the past several decades has examined leader effectiveness increasingly in terms of the leader's interaction with the group and the resulting impact upon outcomes (Stogdill, 1974). Regardless of the researcher's theoretical stance, the definition of leader effectiveness usually focuses on either the leader's behavior or the leader's effect upon group processes or outcomes.

An outgrowth of this situation has been a proliferation of measurement techniques and approaches for assessing the effectiveness of leaders. Typical measurement techniques have included trait ratings, behavior descriptions or perceptions, and personality measures. In addition, researchers often gather data from different sources, including superiors, outside observers, subordinates, or the leaders themselves. Unfortunately, this abundance of methods has resulted in a confusing collection of sometimes uninterpretable and even contradictory findings. The crux of the problem is the attempt to establish the construct validity of the criteria used in leadership research.

One technique for reducing this confusion is the multitrait-multimethod (MT/MM) correlation matrix for examining the convergent and discriminant validity of several measures from several sources (Campbell & Fiske, 1959). Although this technique has been used in several different areas, it has not been used extensively in the leadership area. Yunker and Hunt (1976) used this technique to compare behavioral descriptions of leader effectiveness; and Lawler (1967) and Kavanagh, MacKinney, and Wolins (1971) used it to examine managerial performance ratings from three different sources.

The Kavanagh et al. (1971) study is particularly interesting because it suggests an analysis of variance technique that simplifies the analysis and interpretation of a multitrait-multimethod matrix. The ANOVA technique was applied to 20 ratings from each of three different respondent sources to demonstrate that it provided at least as much information as the cumbersome interpretive requirements of a 60 x 60 matrix.

This study adopted this analysis of variance technique to augment the traditional MT/MM approach and to simplify interpretation of the matrix. Specifically, this study investigated the degree of convergence and divergence among various techniques for evaluating leader effectiveness and utilizing different sources for these evaluations.
Subjects

A 2-week field training exercise for an Army Special Forces Reserve Unit was the setting for this investigation. Data were gathered from 275 reserve unit members from 23 detachments; each detachment contained 10 to 12 enlisted specialists, an executive officer, and a commanding officer. The exercise scenario required each detachment to train another group of reservists in guerrilla-type activities. Data also were gathered from 346 other reservists from the 23 detachments, herein designated as participant observers. Thus, data were collected from three different sources: leaders, group members, and participant observers.

Procedure

Questionnaires were administered at both the beginning and the end of the exercise. All data reported in this report were gathered at the completion of the exercise. Scales administered to the leaders and group members included the Military Leadership Behavior Survey (Downey, 1974), which is similar in methodology and content to the Leader Behavior Description Questionnaire (Stogdill, 1963). The 53-item Military Leadership Behavior Survey provides scores on the following scales: task professionalism, task-oriented consideration, people-oriented consideration, and personal/interpersonal professionalism. In addition, each leader estimated his own performance on a 5-point scale. Also, group members nominated the three peer group members (including the formal leaders) in terms of highest leadership potential.

Scores were derived from questionnaire items about exercise tasks for group and leader ratings of leader effectiveness, group performance, and group relations (Downey, Duffy, & Shiflett, 1975). The questionnaires completed by the participant observers included the same exercise-related items. Although there were differences in the specific items defining the factors across groups (Downey et al., 1975), scores for leader effectiveness, group performance, and group relations were also derived for the participant observers. Table 1 shows the number of items in each scale for the internal and external sources and the number of common items.

RESULTS

The analysis proceeded in three stages. The first step investigated the degree of convergence among measures directed specifically at the leader's behavior. The data necessary for assessing discrimination among all the leader behavior measures were not available for all three sources; hence, the first step in the analysis examined only
the degree of convergence among all the leader effectiveness (direct estimates) measures and from all sources, whether or not a particular variable was available from more than one source. Table 2 shows the correlations among all these measures. The leadership potential measure in Table 2 is the peer evaluation score derived for the formal leaders. The data showed high convergence among all of the measures from the group members, which, as demonstrated by Weissenberg and Kavanagh (1972), is not uncommon for measures of this type. The leader effectiveness measure obtained from the participant observers also showed a fair degree of convergence with the measures from the group members, as well as the leaders' ratings of their own effectiveness. This was the only instance in which the leaders' self-ratings were related to any of the other leader effectiveness measures. The peer evaluations of leadership potential showed little if any relationship to any of the other variables.

Table 1

Item Content of Leader Effectiveness, Group Performance, and Group Relations Measures

<table>
<thead>
<tr>
<th>Variable</th>
<th>Leaders and detachment members</th>
<th>Participant observers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leader effectiveness</td>
<td>6^a</td>
<td>4</td>
</tr>
<tr>
<td>Group performance</td>
<td>17</td>
<td>7</td>
</tr>
<tr>
<td>Group relations</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

^a The leader effectiveness measure for the formal leader was a single item.

The second step of the analysis developed a multitrait-multimethod intercorrelation matrix (Campbell & Fiske, 1959) with the leader effectiveness, group performance, and group relations measures from each of the three sources. This analysis revealed the extent to which the measures were converging or diverging and, consequently, determined whether or not they could be used interchangeably. Table 3 presents these results. Following the guidelines prescribed by Campbell and Fiske, the results showed that for the criterion of convergence—the significance of correlations in the validity diagonals—five of the nine correlations were significant. In general, there was more convergence for the leader effectiveness and group performance measures than for the group relations measure.
Table 2
Intercorrelations of Different Methods of Measuring Leader Effectiveness (N = 23)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Respondent</th>
<th>Variable</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>1. Task professionalism</td>
<td>G</td>
<td>--</td>
<td>97**</td>
<td>86**</td>
<td>89**</td>
<td>84**</td>
<td>27</td>
<td>34</td>
</tr>
<tr>
<td>2. Task-oriented consideration</td>
<td>G</td>
<td>--</td>
<td>90**</td>
<td>89**</td>
<td>84**</td>
<td>27</td>
<td>34</td>
<td>42*</td>
</tr>
<tr>
<td>3. People-oriented consideration</td>
<td>G</td>
<td>--</td>
<td>90**</td>
<td>79**</td>
<td>15</td>
<td>37</td>
<td>44*</td>
<td></td>
</tr>
<tr>
<td>4. Personal/interpersonal professionalism</td>
<td>G</td>
<td>--</td>
<td>85**</td>
<td>32</td>
<td>29</td>
<td>41*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Leader effectiveness</td>
<td>G</td>
<td>--</td>
<td>27</td>
<td>28</td>
<td>44*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Leadership potential</td>
<td>P</td>
<td>--</td>
<td>03</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Leader effectiveness</td>
<td>L</td>
<td>--</td>
<td>51*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Leader effectiveness</td>
<td>O</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Decimal points have been omitted. G = group, P = peers, L = leader, O = participant observers.

*p < .05.

**p < .01.
Table 3

Multitrait-Multimethod Matrix of Measures of Leader Effectiveness, Group Performance, and Group Relations (N = 23)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Detachment</th>
<th>Leader</th>
<th>Participant observers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D1 D2 D3 L1 L2 L3 P1 P2 P3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Leader effectiveness</td>
<td>D1</td>
<td>L1</td>
<td>P1</td>
</tr>
<tr>
<td>2. Group performance</td>
<td>D2</td>
<td>L2</td>
<td>P2</td>
</tr>
<tr>
<td>3. Group relations</td>
<td>D3</td>
<td>L3</td>
<td>P3</td>
</tr>
</tbody>
</table>

*p < .05.
**p < .01.
For the criterion of divergence, three of the validity diagonal values exceeded the respective column and row values, but the remaining six diagonal terms demonstrated one or more reversals. Even less supportive results were evident for the third criterion, requiring higher values in the validity diagonals than in the hetero-trait/mono-method triangles. Only two of the nine values exceeded both of their row and column comparison terms.

An evaluation of the fourth criterion, requiring the same pattern between variables in the hetero-trait/mono-method section as in the hetero-trait/hetero-method section, suggested that not only were the different sources not converging on the same trait but also, in fact, may have been converging across traits.

The third step in the analysis was based on the analysis of variance model described by Kavanagh et al. (1971). In brief, the model is a three-way factorial design from which four sources of variance can be derived to capture the essence of a multitrait-multimethod matrix. These are (a) leader variance, indicating convergent validity; (b) leader by trait variance, indicating discriminant validity; (c) leader by source variance, indicating method/source bias; and (d) error (Kavanagh et al., 1971). In addition, the amount of variance accounted for by each source can be estimated to facilitate relative comparisons among the variance components; this is not feasible in an analysis restricted solely to the correlations in a multitrait-multimethod matrix.

The analysis of variance in Table 4 was derived from the correlations in Table 3. The results indicate fairly strong agreement on the ordering of leaders across the three sources and the three effectiveness measures ($F(22, 88) = 9.06, p < .01$). However, there was no evidence for ordering of leaders on different effectiveness measures (discriminant validity) ($F(44, 88) = 1.49$, N.S.). There was some limited evidence that source/method halo bias was present ($F(44, 88) = 2.00, p < .05$), although it was probably not enough to be of major practical concern. In short, these results support the interpretations made using the traditional multitrait-multimethod approach—that is, there was convergent validity in the data and essentially no evidence for discriminant validity. But the ANOVA results extended the MT/MM results by indicating both the amount of source/method bias and the relatively large amount of error variance involved.

**DISCUSSION**

The data indicated a substantial amount of agreement on the overall ordering of leaders when using a variety of techniques to evaluate leader effectiveness. Further, when leader behavior and group outcome measures were included, there was relatively little differentiation between these types of measures, including the influence of source or relative halo bias (Willingham & Jones, 1958). For example, the participant observers did not seem to distinguish among group relations,
group performance, and leader effectiveness ratings; this suggested that all their ratings were based on a more limited set of cues than in the case of group members and leaders who did show more divergence among these measures.

**Table 4**

Analysis of Variance of Correlations (from Table 3)

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Variance component</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leader</td>
<td>22</td>
<td>4.26</td>
<td>9.06**</td>
<td>.42</td>
</tr>
<tr>
<td>Leader x trait</td>
<td>44</td>
<td>0.70</td>
<td>1.49</td>
<td>.08</td>
</tr>
<tr>
<td>Leader x source</td>
<td>44</td>
<td>0.94</td>
<td>2.00*</td>
<td>.16</td>
</tr>
<tr>
<td>Error</td>
<td>88</td>
<td>0.47</td>
<td></td>
<td>.47</td>
</tr>
</tbody>
</table>

*p < .05.

**p < .001.

The one leader effectiveness variable that did not seem to be related to any of the other variables was peer nomination of leadership potential. This might be a result of a method problem or a function of the peer nominations actually tapping an entirely different construct. Possibly, the respondents closely followed the instructions to nominate individuals for leadership potential; in so doing, they included aspects of behavior and performance that were untapped by the other "here-and-now" measures.

The findings from this study point to the potential for the interchangeability of estimates of leader effectiveness derived from different methods and/or sources; however, the findings also point to a large amount of error variance. Since most research approaches use criteria from only one source or method because of convenience or circumstance, comparisons of results across studies can be made, but only with appropriate caution. This is because, as indicated in this report, criteria are only moderately comparable, and the degree of convergence leaves much of the variance unaccounted for in each of the measures. The limited nature of these data—in size, measures, and type of sample—does not permit a definitive statement about the generality of the results; nevertheless, the data point to the need to investigate further the nature of the relationships between the commonly used criteria of leader effectiveness. Furthermore, the data presented here suggest that different sources or methods may be subject to different types of bias or insensitivity; the data also suggest that several different
methods and sources within the same study may be essential to adequately tap all aspects of leader performance, as well as to make specific findings interpretable in the context of other studies. As Dunnette (1963a, 1963b) suggested, the maxim for leader effectiveness measures appears to be "the more, the better."

The results of this study have added to a growing body of literature (Mitchell, Larson, & Green, 1977; Rush, Thomas, & Lord, 1977; and Lord, Binning, & Thomas, 1977) documenting the importance of rater behavior and context or situational cues regarding perceptions of leader behavior and leadership effectiveness.
REFERENCES


Downey, R. G., Duffy, P. J., & Shiflett, S. Criterion Performance Measures of Leadership and Unit Effectiveness in Small Combat Units. ARI Research Memorandum 75-9, July 1975.


Stogdill, R. M. Manual for the Leader Behavior Description Questionnaire--Form XII. Columbus, Ohio: Bureau of Business Research, Ohio State University, 1963.


