Article Title: Fiedler's Contingency Theory of Leadership: An application of the meta-analysis procedures of Schmidt and Hunter

Name: Mindy Bergman

Pickup at Evans

4235 TAMU
College Station, TX 77843
Fiedler’s Contingency Theory of Leadership: An Application of the Meta-Analysis Procedures of Schmidt and Hunter

Lawrence H. Peters
Department of Administrative Sciences
Southern Illinois University

Darrell D. Hartke
Southern Illinois University

John T. Pohlmann
Department of Guidance and Educational Psychology
Southern Illinois University

The present review was designed to apply the Schmidt and Hunter meta-analysis procedures to the available literature on Fiedler’s Contingency Theory of Leadership. In the present instance, this involved the quantification of the variance in correlations between leader style and performance that can be explained by sampling error. To the extent that the variance in these results across studies can be explained by sampling error, moderator variables, including situational favorability, would be unnecessary theoretical constructions. To the extent that the variance in study results cannot be explained by sampling error, the possibility of one or more moderators exists. Separate meta-analyses were conducted for those studies that led to the specification of the Contingency Theory and those conducted specifically to test the theory. Results suggested that the Contingency Theory was appropriately induced from the studies on which it was based. With regard to those studies conducted specifically to test the Contingency Theory, however, less supportive evidence resulted. Although the theory was supported for all octants except Octant II within the lab data set, only four of the eight octants produced supportive evidence within the field data set. In these instances, the evidence suggests that additional variables need to be specified to account more fully for the variance within those octants that cannot be explained by sampling error alone.

Since its introduction nearly two decades ago, Fiedler’s Contingency Theory of Leadership has been a center of controversy. Although a large number of studies have accumulated, not all have fully supported the theory. Indeed, one of the sources of controversy has been the interpretation of the amount of support this theory has received. Although Fiedler (1971b, 1978) interpreted the accumulated evidence as support for his theory, others (e.g., Ashour, 1973; Graen, Alvares, Orris, & Martella, 1970) interpreted much of these same data as being nonsupportive. Recently, Strube and Garcia (1981) attempted to address this issue by applying procedures that quantify results from previous literature in such a fashion that those results can be integrated and interpreted. Strube and Garcia (1981), therefore, used one type of meta-analysis procedure for reviewing previous literature in a given area. The purpose of the present article is to extend the work of Strube and Garcia (1981) by using a different type of meta-analysis, one aimed at reaching a judgment about the meaningfulness of the existing variance in observed effects across studies. Specifically, the present study uses the meta-analytic procedures developed by Schmidt and Hunter and their colleagues (see, in particular, Hunter, Schmidt, & Jackson, 1982).

Fiedler’s Contingency Theory (Fiedler, 1971b, 1978) suggests that leadership effectiveness is a function of the interaction between the leader and the leadership situation. According to this theory, leaders can be characterized, on the basis of their scores on the Least Preferred Co-Worker (LPC) scale, as being either task oriented or relationship oriented. Favorable leadership situations are those in which leader-member relations, task structure, and position power combine to produce situations in which leadership is effective. Favorable leadership situations are said to be most likely to occur when leader-member relations are positive, the task structure is high, and the leader possesses high levels of position power. Conversely, unfavorable leadership situations occur when leader-member relations are negative, the task structure is low, and the leader lacks position power. Favorable leadership situations are those in which a leader is able to use his or her position power to influence follower behavior in a manner consistent with the leader’s goals. Unfavorable leadership situations are those in which the leader is unable to use his or her position power to influence follower behavior in a manner consistent with the leader’s goals.

Regardless of the data collection problem that has plagued research on the Contingency Theory. Nearly every study that has been conducted with small sample sizes has suggested that the sample sizes may be too small to adequately determine the relationships among the various factors. In recent years, however, a number of studies have been conducted with larger sample sizes. These studies have generally supported earlier research, and the results have been consistent across a wide range of samples. Therefore, the results of these limited sample studies do not appear to be specific to a particular sample. Also, it is important to note that the results of these limited sample studies have not been consistent across all samples. This inconsistency is not surprising given the nature of the correlations between these variables and the performance, the key variable by the theory, were found to be significant.
being either task oriented or person oriented, and leadership situations are said to vary in their degrees of favorableness. Situational favorableness is defined along three dichotomized situational factors (leader-member relations, task structure, and position power), which when combined define eight specific leadership situations, or octants, that differ in their overall favorability. A situation marked by good leader-member relations, high task structure, and strong position power is considered the most favorable leadership situation, whereas a situation marked by poor leader-member relations, low structure, and weak power is considered the least favorable situation. All other leadership situations between these two extremes (e.g., poor leader-member relations, high structure, and strong power) naturally reflect less extreme degrees of situational favorableness.

Fiedler's theory predicts that leaders with a task-oriented leadership style will be more effective in either highly favorable (Octants I-III) or highly unfavorable (Octant VIII) leadership situations. Leaders with a person-oriented style are said to be more effective in situations of moderate situational favorableness (Octants IV-VII). This theory was induced from data collected over a number of years. Since the first formal specification of his theory (Fiedler, 1967), additional research aimed explicitly at validating these hypothesized relations has been conducted. Thus, both developmental and validation data sets exist.

Regardless of the data set, one common problem has plagued research on the Contingency Theory. Nearly every study was conducted with small sample sizes. For example, the sample sizes ranged from 5 to 26 for the developmental data set and from 5 to 53 for the validation data set. Median sample sizes were 6 and 12 for the developmental and validation data sets, respectively. As a result of this limited sample size, the statistical power (Cohen, 1969) to detect true effects was minimal in many studies. It therefore was not surprising that only a small proportion of the correlations between leader style and performance, the key relation predicted by the theory, were found to be statistically significant.

Although few statistically significant findings emerged from studies conducted to validate the Contingency Theory, many of the correlations in each octant were in the direction predicted by the theory. That is, the majority of the correlations, as predicted, were negative in the extremely favorable and extremely unfavorable situations and positive in the moderately favorable situations. In an early review, Fiedler (1971b) focused on these latter results as support for his theory. For those studies he considered acceptable, he reported that the median correlations, within octant, were in the predicted direction in six of the seven octants for which data were available. He bolstered his argument by reporting nonparametric statistics that focused solely on the direction of the obtained results. He reported that 34 of the 45 available correlations were in the predicted direction, a result that was significant ($p < .01$) by the binomial test.

Here, then, is the crux of the problem. Because the studies were underpowered in the first place, the use of traditional significance criteria (i.e., the .05 level) cannot be meaningfully applied. What does it mean to argue that few significant findings have emerged when it was likely, by design, that few such findings could emerge? In contrast, it is difficult to have confidence in results that are marked by their consistent failure to provide significant findings. Although the results regarding the direction of the findings are suggestive of the validity of the theory, the use of the binomial test provides, at best, only indirect and incomplete support for it. Such tests are sensitive only to the sign of the correlation and not to the magnitude of those results.

Strube and Garcia Meta-Analysis

Recently, Strube and Garcia (1981) attempted to resolve this issue by applying a procedure that uses the significance levels of individual studies to estimate the statistical significance of the final result across the entire set of reviewed studies. They used two related techniques for this purpose. Their approach, therefore, represents the use of two of the several quantitative methods for conducting what Glass (1976) termed a meta-analysis of previous literature.
The first technique involved obtaining the exact one-tailed probability for each LPC-performance correlation in their review. These probabilities were then converted into Z scores, where positive Z scores were recorded for results in the predicted direction and negative Z scores were recorded for results in the nonpredicted direction. The obtained Z scores were then combined using Stouffer's formula (see Mosteller & Bush, 1954). The resulting composite Z score, which represents an average value across studies, was then transformed back into the probability that the combined results occurred by chance.

The second technique also used the derived Z scores. In this instance, the Z scores were used to estimate the number of studies that would have to have been conducted and gone unreported in order to bring the significance level for the overall result back to the .05 level. Thus, the second technique estimated the fail-safe number of studies (Cooper, 1979), an estimate of the magnitude of nonsupportive results that would have to be uncovered to negate the results of the overall meta-analysis. The producers used by Strube and Garcia, therefore, integrated the information taken within each study to produce a probability that represents the accumulated information across all studies and a fail-safe N for negating those results.

For the developmental data set, Strube and Garcia (1981) reported supportive results for the combined data set and for all octants except Octant III and Octant VII.1 For the validation data set, supportive results were reported for the combined data set and for two of the octants (Octant I and IV). When the results of the pretheory data set were combined across octants, for example, the computed significance level was $6.44 \times 10^{-12}$ with a fail-safe N of 526. Corresponding results from the validation data set were $6.49 \times 10^{-7}$ and 360. When the validation data set was subdivided into field and lab subsamples, the results again supported the validity of the model when the data were combined across octants ($p < 6.87 \times 10^{-4}$ with a fail-safe N of 58 for the field studies and $p < 1.59 \times 10^{-3}$ with a fail-safe N of 58 for the lab studies). However, for the within-octant analyses, only Octants I and IV produced supportive results within both the lab and field data sets. The results reported by Strube and Garcia (1981), therefore, tended to support the Contingency Theory for both the developmental and validity data sets when all data were combined across octants. Several instances of nonsupport, however, were found within octants.

Schmidt and Hunter
Meta-Analysis

Recently, Schmidt and Hunter and their colleagues (see, e.g., Hunter et al., 1982) reported on a newer set of procedures for conducting meta-analyses that differ from those used by Strube and Garcia (1981) in two important ways. First, as noted by Hunter et al. (1982), procedures that attempt to find the cumulative probability for an entire set of results say nothing about the magnitude of the effect associated with those studies. In this regard, Hunter et al. (1982) noted that “the practical and theoretical implications of an effect depends at least as much on its size as on its existence” (p. 133).

Second, in addition to concentrating on effect sizes, the Schmidt–Hunter procedure differs from other meta-analytical procedures by not taking the variance in observed results across studies at face value. They pointed out that the observed variance in results across studies may arise from two general sources, one of which may be due to the substantive effects of other theoretical factors (i.e., moderator variables) and the other due to the effects of various artifacts. They identified five potential artifacts that impact on the variance of observed results across studies. They include sampling error, study differences in the reliability of measurement, study differences in range restriction, study differences in instrument validity, and computational, typographical, and transcription errors. Hunter et al. (1982) pointed out the necessity of first considering the extent to which the observed variance in results are due to such artifacts. If the observed effects across studies can be accounted for in terms of the operation of only these artifacts, there would be little variance in observed effects to account for in terms of moderator effects. If, however, there is reason to believe that variance above and beyond that attributable to factual sources exists, only then would it be prudent to explore for the existence of variables to explain that residual variance.

Hunter et al. (1982) provided an application of their technique, and Hunter (1981) found that sample size and differential range restriction accounted for 72% of the variance in observed effects! Of these, sampling error was the most important factor, accounting for 71% of the variance due to all three artifacts. Schmidt and Hunter found a very high percentage of the variance in observed results from artifacts in their procedurizations of these procedures, they concluded that no meaningful study differences could be attributed to moderators remained and, therefore, computed a population effect size for studies. In this instance, the estimated effect size was simply the mean of the observed effects, weighted by respective sample sizes.

The Present Review

The present review was designed to replicate the Schmidt–Hunter technique for the Contingency Theory of Leaders. Strube and Garcia review, the present review subdivides the validation studies into field and lab subcategories. Unlike meta-analyses reported by Strube and Garcia (1981), however, the present review focuses on effect sizes and the quantification of the variance in observed effects can be explained by statistical artifact. Sample size was the only informal measure in all studies, sampling error was the only artifact considered.

To the extent that Fiedler's
observed variance in results across studies is due to such artifacts. If the observed variance in effects across studies can be explained in terms of the operation of one or more of these artifacts, there would be little remaining variance in observed effects to be explained in terms of moderator effects. If, in contrast, there is reason to believe that substantial variance above and beyond that due to artifactual sources exists, only then should one explore for the existence of moderator variables to explain that residual variance.

Hunter et al. (1982) provided formulas for estimating the expected variance due to the first three artifacts. On the basis of several applications of their technique, Schmidt and Hunter (1981) found that sampling error and differential range restriction and reliability differences across studies on average accounted for 72% of the variance in observed effects! Of these, sampling error was the most important factor, accounting for 85% of the variance due to all three artifacts. Given that Schmidt and Hunter found a very large percentage of the variance in observed effects to result from artifacts in their previous applications of these procedures, they naturally concluded that no meaningful variance due to moderators remained and, therefore, computed a population effect size estimate across studies. In this instance, the estimated population effect size was simply the average of the observed effects, weighted by their respective sample sizes.

The Present Review

The present review was designed to apply the Schmidt-Hunter technique to Fiedler’s Contingency Theory of Leadership. Like the Strube and Garcia review, the present analysis subdivides the validation studies into both field and lab subsamples. Unlike the recent meta-analysis reported by Strube and Garcia (1981), however, the present meta-analysis focuses on effect sizes and the quantification of the variance in observed effects that can be explained by statistical artifacts. Because sample size was the only information available in all studies, sampling error was the only artifact considered.

To the extent that Fiedler’s Contingency Theory is supported by the data, the following results should be observed:

First, the variance in the combined results, across all studies within a data set, should not be explained as a result of sampling error. If the observed variance across studies can be explained by sampling error, the leadership situation simply could not be a meaningful moderator.

Second, assuming that meaningful variance in these overall results exists, one should find that the observed variance in results within each octant should be zero once sampling error has been considered. In this instance, because Fiedler hypothesized only situational favorableness as the important moderator variable, it should fully account for the variance in effects across studies. If, within each octant, no residual variance above and beyond that due to sampling error is obtained, then support for the moderating effect of situational favorability would be obtained. If, however, variance beyond that due to sampling error is observed, this would imply that other factors, in addition to situational favorability, must be considered in order to explain the results. In this instance, Fiedler’s theory would have to be considered incomplete.

Third, the sign of the effect, within each octant, should be expected to match those hypothesized by Fiedler (1967). To the extent that all within-cell variance is explained by sampling error and the sign of the within-octant effects agree with Fiedler, the literature would completely support the Contingency Theory. If, in contrast, the sign of the effect differs from theoretical expectations, then considerable doubt would be cast on the validity of the theory.

Method

Literature Base

The literature was reviewed to uncover all studies reporting relations between leadership style and performance for interacting groups for which specific octant designations were reported. The present review, therefore, is more limited in scope than the recent review reported by Strube and Garcia (1981), in which studies that extended the theory (e.g., coacting groups) were also considered. The present review is limited to only those specific studies pertinent to the controversy surrounding the validity of the Contingency Theory as originally specified and tested.
The available literature could be classified into two groups within the validation study data set. The first consisted of 20 studies that Fiedler considered to be proper tests of the Contingency Theory. The second included three studies that Fiedler (1971b) claimed were fatally flawed (e.g., Graen, Orris, & Alvaress, 1971) and, thus, not meaningful tests of the theory, and one additional study (Shima, 1968) that Fiedler (1971b) considered appropriate only after reclassifying the results into different octants than originally reported. Results from this latter investigation were used as originally reported by Shima (1968). All investigations, including those that Fiedler considered proper tests and those he considered improper tests, were included in this meta-analysis.

A total of 11 developmental studies and 24 validation studies were identified (see Tables 1 and 2 for specific citations). The 11 developmental studies yielded 37 leader style-performance relations. The 24 validation studies resulted in 100 such relations. Of the validation studies, 12 were lab investigations and were field investigations (yielding 54 and 46 leader style-performance relations, respectively).

In some instances, more than one performance measure was used, and as a consequence, more than one leader style-performance correlation was reported within an octant. In such instances, the correlations were averaged, resulting in the most stable estimate of the leader style-performance relation within that octant (see Hunter et al., 1982). In other instances, the same group of subjects, because they worked on both structured and unstructured tasks, provided results that were entered into each of two octants. In these instances, the data were used as though they were provided by independent samples. Although the analytical procedures that were used assume independence (i.e., this occurred in only one developmental and two validation studies). Because these instances were rare, they were expected to have little influence on the results of the meta-analysis (Hunter et al., 1982).

Four final issues in conducting this review should be noted. First, several instances of conflicting reports of results were observed. For example, several pretheory correlations given in the original studies differed from those reported in the initial discussion of the Contingency Theory (Fiedler, 1967). In all such instances, correlations reported in the original studies were used. Results from one study (Fiedler & Hutchins, 1955, as cited in Fiedler, 1967) could not be directly obtained and were taken from Fiedler (1967). Second, results from one study (Mitchell, 1969) were classified into specific octants in Fiedler (1971b), even though they were not classified into octants in the original report of this study. In fact, Mitchell (1969) did not regard his work as a specific test of the Contingency Theory. It was, therefore, excluded. Third, the data set included Pearson, rank order, and point biserial correlation coefficients. Because these statistics are different forms of the Pearson r, they were considered interchangeable and were used. Finally, two studies (Hardy, 1971; Hardy, Sack, & Harpine, 1973) reported their results in terms of an F statistic. Because their results always involved comparing task-oriented versus person-oriented leaders, they were readily translated into r values based on formulas given in Hunter et al. (1982).

Analytical Procedures

The statistical basis for this meta-analysis rests on the distribution theory of the Pearson correlation coefficient (Hunter et al., 1982, p. 40). The method derives from the limiting distribution of r. In large samples, the distribution of r is approximately normal with an expected value equal to µ, the population correlation, and a variance given by σ^2 = (1 - r^2)/(n - 1). Thus, the distribution of r depends on n and µ.

The goal of the meta-analysis procedure is the estimation of effect parameters. In validity studies of the Fiedler model, the effect size parameter is the population correlation coefficient, and because moderator effects are postulated, the variance of the n is also of interest. The Schmidt and Hunter method estimates µ with a weighted average of the sample correlation coefficients, r = Σi n_i r_i / Σi n_i, where n_i and r_i are, respectively, the sample size and correlation observed in Sample i, the summation being taken over the number of analyses. The variance of r due to sampling error is estimated by S^2 = k(1 - r^2)/N, where k is the number of studies reviewed and N is the total sample size over all k studies. The variance of the r attributable to sampling error and all other sources, to include moderator effects, is estimated by S^2 = Σi n_i (r_i - r)^2/2(n_i - 1). This estimate of the variance of r is a weighted average of the squared deviations of the observed r's and r_i.

If moderator effects are present in a set of correlations, then S^2 should be considerably smaller than S^2. Indeed, previous research by Schmidt and Hunter (Hunter et al., 1982; Schmidt & Hunter, 1961) indicate that the variance due to three artifacts (sampling error and differences across studies in reliabilities and range restriction) accounted for an average of 72% of the variance in study results. Recently, Hunter et al. (1982) suggested a 75% decision rule as a guide to interpreting meta-analysis results involving these three sources of variance. In the present work, however, only sampling error was considered. Because sampling error was shown to account for 85% of the explained variance due to the three artifacts (see Schmidt & Hunter, 1981), it would appear that 60% of the variance due to sampling error alone is less than 60% of the observed variance, one would conclude that a substantial amount of variance in observed effects still remains and a search for moderators would appear warranted. If, in contrast, the variance due to sampling error is equal to or greater than 60% of the observed variance, one would conclude that no meaningful variance in that effect remains and, therefore, a search for moderator variables would be inappropriate.

Results

Results from this review are summarized in Tables 1 and 2 for the pretheory and validation data sets, respectively. Each table summarizes results for each octant and for all data combined. The tables contain the number of correlations observed, number of studies from which derived. The tables also summarize weighted mean correlations and performance along with variance around that mean and due to sampling error. Finally, are summarized with regard to the effect size parameter. The frequency-weighted mean is in the hypothesized direction and the variance due to sampling error is 60% as large as the observed variance in studies.

Because Fiedler induced his results obtained in the pretheoretical data analysis, it is instructive to examine the extent to which the inductive process accurately fits the available data. Results per pretheory data are given in Table 1 and 2. In the first case, the column for all octants combined, the data indicate that sampling error fully accounts for the observed variance and the frequency-weighted mean of results, therefore, are consistent with the conclusion that the leader style-performance is expected to be independent of the values of one or more moderator variable.

When the analyses were repeated on each of the octants for which data was available (all but Octant VI), the hypothesis of the frequency-weighted mean correlation, the hypothesized impact of situation is contradicted by the fact that the results are not significant in the appropriate direction, and there was evidence that the difference in observed variance around those means is not statistically significant, which implies that the explanation in terms of moderator variables would be inappropriate. The specification of the Contingency Theory (Fiedler, 1967) appears to have failed from the data.

Results for the validation data are given in Table 2. As with the pretheoretical combined results, across octants, the results again suggest that sampling error accounts for the observed variance, and the frequency-weighted mean of r is -0.04, therefore, meaningful moderators might exist in this instance, however, only pa...
The basis for this meta-analysis rests on the theory of the Pearson correlation coefficient (e.g., 1982, p. 40). The method derives from the distribution of \( r \). In large samples, the distribution of \( r \) is approximately normal with an expected value of \( \rho \), the population correlation, and a standard error of \( \sqrt{1 - \rho^2}/\sqrt{n-1} \). Thus, the distribution depends on \( \rho \) and \( n \).

The meta-analysis procedure is the estimate of effect parameters. In validity studies of the \( r \), the size effect parameter is the population, and because moderator effects are the variance of the \( r \) is of interest. The Hunter method estimates \( \rho \) with a weighted sample correlation coefficient, \( r_j = \sum n_j r_j / \sum n_j \), and \( r_j \) are, respectively, the sample size obtained in Sample i, the summation over the number of analyses. The variance of \( r_j \) is estimated by \( \sigma^2 = k(1 - k/N) \) where \( k \) is the number of studies reviewed and \( N \) is the sample size per k studies. The variance attributable to sampling error and all other effects is present in a set of correlations, \( r_j \) be considered smaller than \( \sigma^2 \). Indeed, by Schmidt and Hunter (Hunter et al., 1981), the variance artifacts (sampling error and differences in reliability and range restriction) account for about 75% of the variance in study by Hunter et al. (1982) suggested a 75% guide to interpreting meta-analysis since these three sources of variance. In the absence of any sampling error was considered.

The variance accounted for by the three artifacts (see Hunter et al., 1981) it would appear that 60% (75% x .85) would be useful as a decision about whether the variance in observed conditions can be interpreted as evidence of a moderator effect. With this decision rule, if the variance due to the three artifacts is more than 50% of \( \rho \), then it is possible to take into account the substantial variance observed in observed effects remains and moderators would appear warranted. If, in addition to sampling error is equal to 5% of the observed variance, one would not have very little meaningful variance in that effect before a search for moderator variables 2.

Results

In this review are summarized two for the pretheory and a set of tables, respectively. Each table results for each of the octants and for the data. The tables contain the number of correlations observed and the derived number of studies from which they were derived. The tables also summarize the frequency-weighted mean correlation between LPC and performance along with the observed variance around that mean and the variance due to sampling error. Finally, these results are summarized with regard to whether or not the frequency-weighted mean correlation is in the hypothesized direction and whether the variance due to sampling error is at least 60% as large as the observed variance across studies.

Because Fiedler induced his theory from results obtained in the pretheory data set, it is instructive to examine the extent to which that inductive process accurately portrayed the available data. Results pertinent to the pretheoretical data are given in Table 1. Looking first at the frequency-weighted mean correlation combined, the data indicate that sampling error cannot account for the observed variance around the frequency-weighted mean of -.19. These results, therefore, are consistent with Fiedler’s conclusion that the leader-styler-performance relation should be expected to vary with respect to the values of one or more moderators.

When the analyses were repeated within each of the octants for which data were available (all but Octant VI), the results fully accounted for Fiedler’s induction concerning the hypothesized impact of situational favorableness, the hypothesized moderator variable, on the relationship between leader style and performance. In all cases (see Table 1), the frequency-weighted mean correlations were in the appropriate direction, and in no case was there evidence that meaningful residual variance around those means existed for which explanation in terms of additional moderator variables would be needed. Thus, the specification of the Contingency Theory (Fiedler, 1967) appears to have fully followed from the data.

For the test studies, as expected, the results for all octants combined indicated that the observed variance could not be fully understood in terms of sampling error and, therefore, suggested the operation of one or more moderators. With only two exceptions, situational favorability fully accounted for this residual variance in the manner specified by Fiedler. Results for Octants III and VI indicated meaningful residual variance beyond that which could be explained by sampling error.
error alone. In addition, the frequency-weighted correlation for Octant II was .18 and, therefore, in the direction opposite to that predicted by Fiedler, and the frequency-weighted correlation for Octant V was close to zero.

Results pertinent to the field studies once again suggested, for all cells combined, that one or more moderators might meaningfully explain the variance in correlations across studies. When these analyses were repeated within each octant, nonsupportive results were again found within five cells. In Octants I, III, VI, and VIII (which produce nonsupportive findings, the observed within-octant variance correlations was too large to be meaningful. Further, the frequency-weighted mean correlation in Octant V was close to zero and not positive as indicated. Although the frequency-weighted correlation in the predicted direction within each octant, the variance in observed results was explained by sampling error. The

Note. The validation data set (with octants investigated within each study indicated in the parentheses) included: Cherners, Rice, Sundstrom, & Butler, 1975 (VI, VII); Cherners & Skrypeck, 1972 (I-VIII); Csoka, 1975 (I, III, V, VIII); Csoka & Fiedler, 1972 (I, III, V, VIII); Butterfield, 1968 (III, V, VII, VIII); Fiedler, 1966 (I-VIII); Fiedler, 1971a (II, IV); Fiedler & Hutchins, 1955, cited in Fiedler, 1967 (I, V); Fiedler, O'Brien, & Ilgen, 1969 (IV, VI, VII, VIII); Graen et al., 1971 (I-VIII); Hardy, 1971 (I-VIII); Hardy et al., 1973 (I-VIII); Hill, 1969 (I, III, V, VII, VIII); Hovey, 1974 (IV, VIII); Hunt, 1967 (I, III, V, VII, VIII); McNamara, 1967 (IV, VIII); O'Brien, Fiedler, & Hewett, 1969 (II, IV, VI, VII, VIII); Rice & Cherners, 1973 (I-VIII); Subin, 1972 (III); Sashkin, Taylor, & Tripathi, 1974 (III, VII); Sashkin, 1978 (II); Shiftlet & Nealey, 1972 (III, IV); Shima, 1968 (I, III, IV, and Tatum, 1972 (I, IV, VIII). Fiedler (1966) and Mitchell (1969) reported correlations between two or more performance measures and leader style. In these instances, the mean leader style-performance correlation was used in the meta-analysis.
Table 3

Meta-Analysis Results for the Laboratory Studies in the Validation Set

<table>
<thead>
<tr>
<th>Octants</th>
<th>Summary variables</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
<th>VI</th>
<th>VII</th>
<th>VIII</th>
<th>All octants</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. independent studies</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>7</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>No. correlations</td>
<td>6</td>
<td>7</td>
<td>9</td>
<td>7</td>
<td>4</td>
<td>7</td>
<td>4</td>
<td>8</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td>( \tilde{r} )</td>
<td>-.32</td>
<td>.21</td>
<td>-.11</td>
<td>.27</td>
<td>.01</td>
<td>.33</td>
<td>.25</td>
<td>-.33</td>
<td>.16</td>
<td></td>
</tr>
<tr>
<td>S^2</td>
<td>.081</td>
<td>.059</td>
<td>.125</td>
<td>.059</td>
<td>.113</td>
<td>.162</td>
<td>.040</td>
<td>.092</td>
<td>.158</td>
<td></td>
</tr>
<tr>
<td>S^2</td>
<td>.067</td>
<td>.067</td>
<td>.054</td>
<td>.065</td>
<td>.065</td>
<td>.138</td>
<td>.065</td>
<td>.121</td>
<td>.067</td>
<td>.079</td>
</tr>
<tr>
<td>S^2</td>
<td>-.014</td>
<td>-.005</td>
<td>.071</td>
<td>-.006</td>
<td>-.025</td>
<td>.097</td>
<td>-.081</td>
<td>.025</td>
<td>.079</td>
<td></td>
</tr>
<tr>
<td>More than 60% of variance?</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Sign of ( r ) appropriate?</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Note. The laboratory studies in the validation data set (with octants investigated within each study indicated in the parentheses) included: Chemers et al., 1975 (VI, VIII); Chemers & Skrzymek, 1972 (I, II, III, IV, V, VI, VII, VIII); Fiedler, 1966 (I, II, III, IV, V, VI, VII, VIII); Fiedler, 1971a (II, IV); Graen et al., 1971 (I, II, III, IV, V, VI, VII, VIII); Hardy, 1971 (I, II, IV, IV); Hardy et al., 1973 (II, IV, VI, VII, VIII); Hovey, 1974 (IV, VIII); Rice & Chemers, 1973 (VI, VIII); Sashkin, 1971 (II, IV); Shifflet & Nealey, 1972 (III, IV); and Shima, 1968 (I, III). Fiedler (1966) and Mitchell (1969) reported correlations between two or more performance measures and leader style. In these instances, the mean leader style-performance correlation was used in the meta-analysis.

again found within five cells. In this data set, Octants I, III, VI, and VIII continued to produce nonsupportive findings. In all cases, the observed within-octant variance in the correlations was too large to be explained by sampling error. Further, the frequency-weighted mean correlation in Octant VI was close to zero and not positive as hypothesized. Although the frequency-weighted mean was in the predicted direction within Octant IV, the variance in observed results could not be explained by sampling error. These findings, naturally, imply that additional factors, yet unspecified, might explain the variance in correlation within this cell.

Discussion

Contrary to Strube and Garcia's (1981) conclusion that "the model as a whole was overwhelmingly supported" (p. 316), the results of the present meta-analysis present a mixed set of findings concerning the Contingency Theory. On the positive side, results

Table 4

Meta-Analysis Results for the Field Studies in the Validation Data Set

<table>
<thead>
<tr>
<th>Octants</th>
<th>Summary variables</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
<th>VI</th>
<th>VII</th>
<th>VIII</th>
<th>All octants</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. independent studies</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>12</td>
</tr>
<tr>
<td>No. correlations</td>
<td>8</td>
<td>4</td>
<td>9</td>
<td>4</td>
<td>6</td>
<td>3</td>
<td>5</td>
<td>7</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>( \tilde{r} )</td>
<td>-.42</td>
<td>-.33</td>
<td>.28</td>
<td>.45</td>
<td>-.01</td>
<td>.13</td>
<td>-.50</td>
<td>.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S^2</td>
<td>.122</td>
<td>.32</td>
<td>.221</td>
<td>.173</td>
<td>.074</td>
<td>.207</td>
<td>.143</td>
<td>.122</td>
<td>.254</td>
<td></td>
</tr>
<tr>
<td>S^2</td>
<td>.045</td>
<td>.038</td>
<td>.065</td>
<td>.052</td>
<td>.048</td>
<td>.115</td>
<td>.138</td>
<td>.040</td>
<td>.071</td>
<td></td>
</tr>
<tr>
<td>S^2</td>
<td>.077</td>
<td>-.007</td>
<td>.156</td>
<td>.121</td>
<td>.026</td>
<td>.092</td>
<td>.005</td>
<td>.082</td>
<td>.183</td>
<td></td>
</tr>
<tr>
<td>More than 60% of variance?</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Sign of ( r ) appropriate?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Note. The field studies in the validation data set (with octants investigated within each study indicated in the parentheses) included: Butterfield, 1968 (III, VIII); Csoka, 1975 (I, III, V, VIII); Csoka & Fiedler, 1972 (I, III, V, VIII); Fiedler & Hutchins, 1955, cited in Fiedler, 1967 (I, V); McNamara, 1967 (IV, VIII); Fiedler, O’Brien & Ilgen, 1969 (II, IV, VI, VIII); Hill, 1969 (II, III, VI, VII); Hunt, 1967 (I, III, V, VIII); McNamara, 1967 (IV, VIII); O’Brien, Fiedler, & Hewett, 1969 (II, IV, VI, VIII); Sashkin, Taylor, & Tripathi, 1974 (III, VII); Schneier, 1978 (II); and Ternes, 1972 (I, IV).
from the developmental data set fully support the theory on which they were based. The results suggested that meaningful residual variance exists within the combined data set, and that situational favorableness can fully account for that variance. Additional findings from the validation data set also suggested that the theory is supported within three cells (means in the predicted direction and no meaningful residual variance) and partially supported within the other cells (means in the predicted direction but unexplained residual variance after accounting for sampling error). Taken at face value, these results are impressive. They reflect a consistency of findings regarding a set of complex predictions based on research conducted over a two-decade period.

On the negative side, however, results from both lab and field studies cast doubt on the model as originally specified. In general, it would appear that much of the empirical support for this theory is based on lab findings. Indeed, with few exceptions, these findings provide support for Fiedler's theory. Most damaging were the results for Octant II. Instead of observing a negative leader style-performance relation within Octant II, a positive relation (.21) was observed. Further, no residual variance around this relation was observed. We cannot dismiss this finding as quickly as others who observed the same result (e.g., Fiedler, 1978; Strube & Garcia, 1981) and preferred to attribute it to the difficulty of producing Octant II situational conditions in the lab. We find it difficult to accept this argument given that (a) there is no a priori reason to believe that Octant II conditions should be more difficult to create in the lab than any of the other leadership situations, (b) the authors of the several studies in which situational favorableness was manipulated to produce an Octant II setting said nothing about the difficulty in doing so, and (c) reviewers of those papers seemingly saw nothing contrived about those experimental settings (more so than in the manipulation of any other octant), which prevented them from recommending those papers for publication. Thus, we can only conclude that the theory is inaccurate with regard to this leadership condition, at least as it is created within lab settings.

The results from the field studies fall short of providing full support for the Contingency Theory. Because the theory ultimately must be useful within field settings, it is crucial that supportive findings emerge in such settings. In this instance, meaningful residual variance, above and beyond that due to artifacts, was observed in five of the eight octants (Octants I, III, IV, VI, and VIII). Such results suggest that the theory is incomplete in specifying important moderating factors. The near-zero leader style-performance association in Octant VI is contrary to expectations, and therefore must also be accounted for in additional theoretical specifications of the Contingency Theory.

It should be recalled that the validation data set included three studies (Butterfield, 1968; Graen et al., 1971; McNamara, 1967) that Fiedler contended were flawed and one study (Shima, 1968) in which he contended that the results were improperly classified into octants. To examine the support for this theory more fully, the meta-analysis was repeated without the results from these four studies. We reasoned that this more limited data set should provide the greatest opportunity to find supportive results for the validity of the theory, and therefore cast the greatest doubt concerning its validity if nonsupportive results were obtained.

The reanalysis yielded highly similar findings. With regard to all validation studies in this reduced literature base, only the results for Octant VIII changed. In this instance,
sampling error accounted for the variance in correlations. Thus, four as opposed to five octants produced meaningful variance explained by sampling error. Within the lab data set, all octants produced supportive findings with the exception of Octant II. As before, the leader style–performance correlation was positive ($r = .26$) and, therefore, in the opposite direction. Finally, the reanalysis of the field studies only resulted in one change. That again involved Octant VIII and indicated that sampling error explained the variance in correlations. Results for the field studies, therefore, continue to suggest that sampling error cannot account for variance in the correlations in four of the eight octants.

One last issue involves the small number of within-octant correlations on which the present meta-analysis is based. The number of correlations, per octant, average from 5.28 in Table 1 to 12.50 in Table 2. Because so few results might be randomly unrepresentative of the larger population of potential findings, they raise the possibility of second-order sampling error. Like ordinary sampling error, second-order sampling error would be expected to affect standard deviations more so than means and, therefore, is relevant to our interpretation of the extent to which the within-octant variance in correlations can be explained by differences in sample sizes across studies. Because the number of within-octant studies was small, a second-order meta-analysis was conducted.

The second-order meta-analysis examined the extent to which the variance in the eight computed within-octant correlations could be explained by the varying, and small, number of study correlations on which they are based. This, obviously, is analogous to the first-order meta-analyses reported previously. In keeping with the procedures for conducting a first-order meta-analysis, each within-octant variance estimate, $S^2$, and $S^2$, was weighted by the number of cases on which it was based and pooled to produce within-octant estimates of $S^2$ and $S^2$. This ensures that the more stable estimates receive greater weight than less stable estimates. In further keeping with the first-order meta-analysis procedures, the 60% rule was once again applied. In this case, if second-order sampling error accounted for as much as 60% of the observed variance in within-octant correlations, one would conclude that the octant designation fully accounted for the study results. If second-order sampling error cannot account for as much as 60% of this variance, one would conclude that additional variables, beyond those specified by the Contingency Theory, would need to be specified.

The second-order meta-analysis was conducted on the pretheory and validation data sets as well as the laboratory and field studies within the validation data set. Results for this analysis are given in Table 5. Consistent with the results of the first-order meta-analysis, these results indicate that second-order sampling error accounts for the variance in the within-octant correlations within the pretheory data, but fails to do so within the overall validation data set or the field studies within this data set. Within the laboratory data set, results suggest that the variance in within-octant correlations can be explained by sampling error and, therefore, support Fiedler's model.

The second-order meta-analyses were repeated within the validation data set using only those studies that Fiedler contended were appropriate tests of the Contingency Theory. Similar findings emerged. Second-order sampling error was able to account for the variance only within the lab studies. Taken as a whole, the present first-order and second-order meta-analyses lead to a generally positive conclusion regarding the validity of the Contingency Theory, similar to that reached by Strube and Garcia (1981). We concur with them in concluding that considerable evidence exists in support of the Contingency Theory. Indeed, only one result (i.e., the mean correlation in Octant II of the validation data set) was contrary to theoretical expectations. We do not want this point to be overlooked. Nonetheless, we differ from Strube and Garcia’s (1981) conclusion in one important aspect. We believe that to rest on the supportive evidence belies the data that suggest the Contingency Theory, as stated, is incomplete. Even if one were to base a conclusion solely on those studies that Fiedler...
Table 5  
Percent of Validity Coefficient Variance  
Due to Sampling Error From the  
Second-Order Meta-Analysis

<table>
<thead>
<tr>
<th>Data set</th>
<th>Literature base</th>
<th>Restricted literature base</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-theory data set</td>
<td>&gt;100%</td>
<td>IDB</td>
</tr>
<tr>
<td>Validation data set</td>
<td>All studies 51%</td>
<td>56%</td>
</tr>
<tr>
<td></td>
<td>Lab studies 72%</td>
<td>&gt;100%</td>
</tr>
<tr>
<td></td>
<td>Field studies 42%</td>
<td>53%</td>
</tr>
</tbody>
</table>

Note. IDB = Identical data base for the pre-theory studies.
*Excludes Butterfield, 1968; Graen et al., 1971; McNamara, 1967; and Shima, 1968.

did not dispute, a data set for which negative findings should be considered crucial to the theory, the present results suggest that further theoretical construction is needed.

References
Assumed similarity measures as predictiveness. *Journal of Abnormal and Social Psychology*, 49, 381-388.


Received January 31, 1984
Revision received September 27, 1984.