Small Group Research

http://sgr.sagepub.com

A Meta-Analysis of Cohesion and Performance: Effects of Level of Analysis and Task Interdependence

Stanley M. Gully, Dennis J. Devine and David J. Whitney

Small Group Research 1995; 26; 497

DOI: 10.1177/1046496495264003

This study extends previous work reviewing the cohesion-performance relationship by using meta-analytic techniques to assess the effects of level of analysis and task interdependence on the cohesion-performance relationship. A total of 51 effect sizes from 46 empirical studies were obtained for the meta-analytic integration. Results suggest that level of analysis and task interdependence moderate the cohesion-performance relationship. Implications of the findings for future research on group cohesion and performance are discussed.

A META-ANALYSIS OF COHESION AND PERFORMANCE
Effects of Level of Analysis and Task Interdependence

STANLEY M. GULLY
DENNIS J. DEVINE
DAVID J. WHITNEY
Michigan State University

Researchers have frequently considered cohesion to be an important component of group process and performance (e.g., Shaw, 1971; Stogdill, 1972). However, in spite of strong theoretical expectations that cohesion and performance should be positively related (e.g., Cartwright, 1968; Cattell, 1948; Davis, 1969; Shaw, 1971), research has generated conflicting evidence. For example, in a review of the relationship between group drive, group cohesion, and group productivity, Stogdill (1972) found no clear relationship between cohesion and productivity.

AUTHORS' NOTE: *The authors listed contributed equally to the preparation of this article. We would like to thank Neal Schmitt, Jose Cortina, John Hunter, Jean Phillips, Albert Cota, and three anonymous reviewers for their thoughtful input and feedback in response to earlier versions of this article. Requests for reprints should be sent to Stanley M. Gully, Department of Psychology, 129 Psych Research Building, Michigan State University, E. Lansing, MI 48824-1117.*

SMALL GROUP RESEARCH, Vol. 26 No. 4, November 1995 497-520

Recently, Evans and Dion (1991) used meta-analytic techniques to integrate the cohesion literature. They examined 18 coefficients from 16 studies relating various measures of cohesion to various measures of performance. Their findings suggest that cohesion and performance are positively associated, but there are a number of conceptual and methodological issues of concern.

On the conceptual side, no attempt was made to identify potential moderators, although sampling error accounted for only 64% of the variance in obtained effect sizes. Hunter and Schmidt (1990) suggested that when sampling error accounts for less than 75% of the variance in effect sizes, and theory is available, examination of moderators should be pursued. Also, Evans and Dion (1991) appropriately limited their analysis to cohesion-performance relationships at the group level. Although cohesion is a group-level construct, past research has often measured individual perceptions of cohesion and correlated them with individual performance. These studies were excluded by Evans and Dion, but they may have contributed to the current confusion in the cohesion literature (e.g., Mudrack, 1989a). Therefore, we included the individual-level studies in the present investigation.

There are also several concerns with the methodology employed by Evans and Dion (1991). First, their meta-analysis identified only 18 effect sizes in the literature. Of these, a number of discrepancies exist between the reported effect sizes used by Evans and Dion and those used in the present study (see Method section). Second, the large confidence interval obtained by Evans and Dion (.085 to .643) offers little information beyond conclusions of previous qualitative reviews.

A second and more comprehensive meta-analysis was conducted by Mullen and Copper (1994). In their study, they examined several variables that could affect the relationship between cohesion and performance, including interaction requirement, type of cohesion, study type, and time. Overall, the cohesion-performance relationship was small but significant (average r = .248), stronger for correlational studies than for experimental studies (r = .25 vs. .23, respectively), and stronger for real groups than for artificial groups (r = .27 vs. .16, respectively). Their study did not, however, find a significant effect on the cohesion-performance relationship for interaction requirement, but it did find a difference in effect size for group type (e.g., sports, military, and other). Mullen and Copper also posited, based on cumulation of effect sizes over time, that the causal direction of the cohesion-performance relationship is likely to run from performance to cohesion rather than the reverse.

Once again, key conceptual and methodological issues warrant a cautious interpretation of some of their findings. First, Mullen and Copper (1994) performed an analysis of cross-lagged relationships over time by aggregating across Time 1 (T1) and Time 2 (T2) measurements without considering the duration of the time interval. For example, in Williams and Hacker (1982), the T1 measurement of cohesion occurred prior to the completion of two regular season games during a season that lasted 12 weeks with 12 to 14 games, so that the T1 measurement occurred approximately 2 weeks into the study. In contrast, Bakeman and Helmreich (1975) assessed cohesion at T1 over an unspecified time period, Greene (1989) assessed T1 cohesion some time between 5 and 15 months beyond group formation, and Dorfman and Stephan (1984) made their T1 measurement 6 weeks into their study.

The problem with aggregating effect sizes across time is exacerbated for the T2 measurements. The T2 measurement of cohesion occurred 12 weeks into Dorfman and Stephan's (1984) study, after an unspecified amount of time in Bakeman and Helmreich's (1975) study, and more than 35 weeks into Greene's (1989) study. If it is true that the cohesion-performance relationship varies as a function of time, then these discrepant time intervals severely limit the usefulness of an analysis that combines across varying time periods. To obtain meaningful results, it is necessary to measure the cohesion-performance effect across constant time intervals, with sufficient sample size to justify aggregation. Unfortunately, there are too few studies within a given time period to perform an analysis using cross-lagged relationships (Gully, Whitney, & Devine, 1993).

A second and perhaps more important problem with Mullen and Copper's (1994) study is that the findings are confounded by levels of analysis. Failure to consider levels of analysis when analyzing the effect sizes can lead to fallacious conclusions (see discussion below). Thus it is not clear whether the obtained differences in effect size are due to differences in the way cohesion was conceptualized and operationalized (individual vs. group) or some other moderating variable.

Despite the Evans and Dion (1991) and the Mullen and Copper (1994) studies, there remains a large degree of confusion regarding the relationship between cohesion and performance. The present study first extends the work of previous researchers by incorporating a levels-of-analysis perspective. A second extension of previous work is a reexamination of the effect of task interdependence (i.e., Mullen and Copper's interaction requirement) after controlling for levels of analysis.

A LEVELS OF ANALYSIS PERSPECTIVE

One of the shortcomings of the cohesion literature has been the failure to explicitly address levels-of-analysis issues. Increasingly, organizational researchers have emphasized the importance of explicitly stating the appropriate level of analysis for each construct of interest (Ostroff, 1993; Roberts, Hulin, & Rousseau, 1978; Rousseau, 1985).

Level of analysis refers to the unit to which the data are assigned for hypothesis testing and statistical analyses (Rousseau, 1985). There are many possible levels of analysis, including individual, dyad, group, and organization. It is important to note that the level of analysis can be different from the level of the construct of interest. For example, one can measure individual attraction to the group and correlate this with individual performance. The level of analysis is the individual in this case. Alternatively, one can use the mean of individual attractions-to-the-group ratings to represent group cohesion and correlate this with group performance. In this instance, the level of analysis is the group.

Thorndike (1939) demonstrated that when the level of analysis does not match the theoretical level of interest, researchers may draw incorrect or fallacious conclusions regarding the focal level of interest. Subsequent work by Robinson (1950), Roberts et al.

(1978), Ostroff (1993), and Klein, Dansereau, and Hall (1994) has provided further demonstration that results for a different level of analysis cannot automatically be generalized to the focal level of interest. A mismatch between focal level and level of analysis is termed *misspecification*, or the fallacy of the wrong level (Roberts et al., 1978).

Unfortunately, researchers have often operationalized the cohesion construct at the individual level of analysis and generalized findings to the group level. Alternatively, in an attempt to assess cohesion as a group-level construct, many researchers have measured cohesion by aggregating individual responses. When aggregating data to represent a higher level construct, it is assumed that the aggregated variable is isomorphic in function with the construct at a higher level (Mossholder & Bedeian, 1983; Rousseau, 1985).

Although aggregated individual data may be an appropriate way to assess group cohesion, it is first necessary to examine the degree of consensus at the individual level. Only when there exists an acceptable degree of consensus can individual data be aggregated to form a group-level construct. Indexes exist to measure the extent of consensus prior to aggregation (James, Demaree, & Wolf, 1984; Kozlowski & Hattrup, 1992). Failure to consider consensus when aggregating data at one level to represent a higher level of analysis may result in aggregation bias, a class of errors in which an observed relationship is an artifact of the data combination method (James, 1982). Unfortunately, an examination of the cohesion literature reveals that individual agreement concerning group cohesion has been rarely assessed prior to aggregation to the group level.

Ostroff (1993) pointed out that although group-level correlations can be higher, lower, or unchanged in comparison to individual-level correlations, relationships at higher levels of analyses are often observed to be stronger than at the individual level of analysis. These changes in correlation may be due to statistical artifacts, elimination of error variances, biased estimates, or a reflection of an actual difference between individual-level and higher level constructs. Thus there exist methodological reasons that suggest a stronger cohesion-performance relationship will be observed when cohesion is operationalized as a group-level construct, particularly when aggregation is performed without consideration for consensus.

Theory also predicts that cohesion operationalized at the group level of analysis will be more highly related to performance than when operationalized at the individual level. This statement is based, in part, on Indik's (1968) theoretical argument that variables at the same level of analysis should be more highly related than variables at different theoretical levels. Because group performance is a group-level variable, the cohesion-performance relationship is expected to be stronger when cohesion is operationalized at the group level of analysis. Failure to explicitly consider levels-of-analysis issues in previous studies may have contributed to the inconsistent findings obtained in the literature.

TASK INTERDEPENDENCE

A second important construct often overlooked by researchers that may have contributed to the inconsistent findings in the cohesion-performance literature is task interdependence. Some tasks, such as flying a passenger jet, performing surgery, directing military operations, or playing a game of basketball, require high levels of interaction among group members (Sundstrom, De Meuse, & Futrell, 1990). For such tasks, one might expect group cohesion to strongly affect constructs such as group coordination, cooperation, and communication, as well as group performance. Cohesion should also affect individual motivational factors such as direction, persistence, and intensity, which will, in turn, affect group processes. Thus, in highly interdependent tasks, cohesion operates to affect individual motivational factors, group processes, and group outcomes. The result should be a strong cohesion-performance relationship for interdependent tasks (Widmeyer, Brawley, & Carron, 1992).

Other tasks, such as golf, bowling, and production-line manufacturing, are primarily individual tasks and do not necessitate group interaction. In these so-called group tasks, the individual usually knows what to do and how to do it, and there is little need for the group to coordinate, communicate, or cooperate. For tasks such as these, one might expect that group cohesion will be less strongly related to subsequent performance outcomes because cohesion only operates to influence performance through individual motivational processes. Thus cohesion should be less strongly related to group performance on tasks that require limited group member interaction than on tasks that require large amounts of group interaction (Widmeyer et al., 1992). Of note, Mullen and Copper (1994) did not find a significant effect for a construct similar to task interdependence (i.e., interaction requirement). This is perhaps attributable to the confounding of the levels of analysis.

To summarize, the cohesion-performance literature has tended to overlook the issues of levels of analysis and task interdependence. When studies are conducted on groups with low task interdependence or when they use individual-level measures, we expect group cohesion to be less strongly related to performance than when member interdependence is high or when group-level measures are employed. The current study is designed to address these issues.

METHOD

Three approaches were used to identify studies for the current meta-analysis. First, computer searches were made of the I/O Reference, Infotrac, and PsychLit databases. Second, manual searches were conducted using the references from several of the more recent or comprehensive empirical and theoretical review papers (e.g., Bettenhausen, 1991; Carron, 1982, 1988; Evans & Dion, 1991; Levine & Moreland, 1990; Lott & Lott, 1965; Mudrack, 1989a, 1989b; Stogdill, 1972; Zander, 1979). Manual searches were also conducted for the following journals for the last 10 years: *Group and Organization Studies*, *Journal of Applied Psychology*, *Journal of Social Psychology*, *Journal of Sport Psychology*, *Organizational Behavior and Human Decision Processes*, *Organizational Dynamics*, and *Small Group Research*.

The criterion for initial selection in these search efforts was mention of the terms *performance*, *effectiveness*, or *productivity* in the articles, as well as either *cohesion* or *cohesiveness*. Of the hundreds of original studies searched, 116 empirical studies remained after unobtainable, nonempirical, and/or redundant studies had been removed. The unobtainable studies primarily consisted of unpublished theses, dissertations, and reports, and they comprised less than 10% of the total body of literature identified. An effort was made to procure all available empirical studies.

A total of 70 of the 116 empirical studies were discarded from the analysis during feasibility examination. Studies were discarded if they (a) failed to relate cohesion and performance with a usable statistic or (b) failed to report the means and standard deviations necessary for formula transformations. A total of 46 empirical studies with 51 effect sizes remained after feasibility analysis for inclusion in the meta-analysis.

We read and coded each of the 46 remaining studies for task interdependence, level of analysis, sample size, reliabilities of independent and dependent variables, and effect sizes. Sample sizes were taken directly from Methods and Results sections of the studies, with special reference to the level of analysis reported. Reliabilities of measurement instruments were used whenever reported. For the dependent variable of win-loss records, a reliability of 1.0 was assigned because it was assumed that no error of measurement existed for this variable. The effect of this decision was to yield a higher average reliability in the dependent variable and less correction for attenuation when making population estimates. The average reliabilities were .836 for measurement of cohesion and .832 for measurement of performance.

Effect sizes were calculated using standard transformation formulas reported by Hunter and Schmidt (1990). The following procedures were used to determine effect sizes: (a) use zero-order correlations, F tests (1 degree of freedom in numerator), t tests, or means and standard deviations whenever present; (b) average across synchronous correlations if multiple correlations across time are reported; (c) use objective performance criteria over subjective criteria whenever possible; and (d) average across multiple objective criteria when multiple criteria are present. When both social and task cohesion were measured, effect sizes were averaged across both measures because there were too few studies currently available to warrant a moderator analysis for type of cohesion. Averaging effect sizes was chosen over reporting each effect size separately in the meta-analysis to avoid artificial inflation of sample sizes.

Task interdependence was determined for each task based on ratings for each of the following three dimensions: communication, coordination, and mutual performance monitoring. These dimensions were derived from conceptual definitions of task interdependence presented in previous research (e.g., Mitchell & Silver, 1990; Saavedra, Earley, & Van Dyne, 1993; Salas, Dickinson, Converse, & Tannenbaum, 1992; Shea & Guzzo, 1987). If a task was coded high for two or more of the three dimensions, it was considered to be high in task interdependence. Conversely, if a task was coded low on two or more of the three dimensions, it was considered to be low in task interdependence. Ratings were made independently by three raters, and pairwise kappa agreement coefficients were computed. The kappas were very high (.85) for all three pairs of raters, reflecting two disagreements per rating pair out of a possible 35 studies coded.

Level of analysis was coded by examining whether the cohesion-performance relationship was tested at the individual or group level. This was determined by examining the sample size of the study, the degrees of freedom in the denominator, and by the author's description of the method of analysis. Pairwise kappa values for the coding of level of analysis were .92, .78, and .70. It should be noted that kappas are indexes of agreement, not reliability (Tinsley & Weiss, 1975), and a single discrepancy has a large effect on kappa when there are a small number of rating categories. Thus the obtained kappas are reasonably high given the number of categories coded for level of analysis.

After formulation of consensus on every coefficient in the current study, Evans and Dion's (1991) study was consulted as a further check on coding accuracy for sample and effect sizes. Consistent with Mullen and Copper (1994), this check resulted in the surprising finding of discrepancies for 10 of the 18 coefficients reviewed in Evans and Dion's meta-analysis. In addition, three studies (4 coefficients) were included in Evans and Dion's study, although we could not agree on a reliable effect size (because of inadequate descriptions of methods used in these studies); these studies were not included in current analyses. Our effect sizes tended to be smaller than those provided by Evans and Dion. Interestingly, our obtained effect sizes also diverged somewhat from those of Mullen and Copper, but not in a systematic manner. Overall, the differences between our effect sizes and those obtained by Evans and Dion and by Mullen and Copper were relatively minor and are possibly attributable to the many judgment calls regarding theory and method that have to be made during meta-analysis (C. R. Evans, personal communication, March, 1994; Wanous, Sullivan, & Malinak, 1989). All discrepancies were noted, and discrepant effect sizes were further reviewed. Careful evaluation of effect size transformation formulas followed, with subsequent recalculations. In spite of these procedures, no changes were deemed necessary.

A total of 46 studies with 51 effect sizes constitute the empirical literature reviewed in this meta-analysis. Table 1 summarizes the final coding of the variables for all studies included in the meta-analysis, along with the reliabilities for the cohesion and performance measures. The complete set of references for the studies included in the meta-analysis is available upon request.

The techniques used to cumulate the findings were those outlined by Hunter and Schmidt (1990). First, a meta-analysis was conducted that included all 51 effect sizes, without regard to moderators (full analysis). In the second analysis, the study by Hoiberg and Pugh (1978, Study 17) was removed because its large sample size (N = 7,923) heavily weighted the results of the meta-analysis and also because it suffered from a serious construct deficiency. Hoiberg and Pugh defined cohesion as how cohesive a person expected to be with new people after a job reassignment.

Also, all studies that used a self-report measure of productivity were removed in this second step because of (a) likelihood of method bias, (b) low validity with objective productivity measures, and (c) low reliabilities. Support for the presence of method bias is found in the corrected weighted mean effect size of the self-report studies (r = .65). The decision to remove studies using self-report measures of productivity resulted in the removal of six effect sizes (10, 26, 27, 32, 33, and 34). The removal of the Hoiberg and Pugh

TABLE 1: Coded Characteristics of the Studies

Number	*Study*	r_{xx}	r_{yy}	*TI*[a]	*Level*[b]	N[c]	*Effect*
1	Bakeman and Helmreich (1975)	.93	—	1	2	10	.645
2	Ball and Carron (1976)	—	1.00	2	2	12	.659
3	Bird (1977)	—	1.00	2	2	8	.790
4	Cohen, Whitmyre, and Funk (1960)	—	—	—	1	16	.289
5	Colarelli and Boos (1992)	.86	0.73	1	2	86	.050
6	Dailey (1978)	.98	0.65	—	1	281	.188
7	Darley, Gross, and Martin (1952)	—	0.85	1	2	13	.455[d]
8	Deep, Bass, and Vaughan (1967)	—	—	1	2	9	–.220
9	Dorfman and Stephan (1984)	—	—	—	1	93	.325
10	Fandt (1991)	.84	—	1	2	115	.240
11	Fiedler (1954)	—	—	1	2	18	–.230
12	Gekoski (1952)	.84	0.87	1	2	21	.100[d]
13	George and Bettenhausen (1990)	.85	—	1	2	33	.040
14	Goodacre (1951)	—	—	2	2	12	.720
15	Greene (1989)	.84	0.65	1	2	54	.100
16	Hemphill and Sechrest (1952)	.91	0.80	2	2	80	.360
17	Hoiberg and Pugh (1978)	—	—	—	1	7,923	.080[d]
18	Hoogstraten and Voorst (1978)[e]	.82	—	M	2	16	.720
19	Hoogstraten and Voorst (1978)[e]	.82	—	—	1	91	.141
20	Hoogstraten and Voorst (1978)[e]	.82	—	—	1	126	.197
21	Jaffe and Nebenzahl (1990)	—	—	1	2	20	.175[d]
22	Keller (1986)	.77	0.74	1	2	32	.395
23	Keyton and Springston (1990)	.92	0.87	1	2	35	–.116
24	Klein and Christiansen (1969)	—	—	2	2	35	.248
25	Landers, Wilkinson, Hatfield, and Barber (1982)	—	1.00	2	2	10	.753
26	Littlepage, Cowart, and Kerr (1989)[f]	—	—	—	1	95	.320
27	Littlepage, Cowart, and Kerr (1989)[f]	—	—	—	1	49	.370
28	Melnick and Chemers (1974)	—	1.00	1	2	21	.063
29	Moos and Speisman (1962)	—	—	2	2	60	.360
30	Norris and Niebuhr (1980)	.68	—	1	2	18	.440
31	Piper, Marrache, Lacroix, Richardsen, and Jones (1983)	—	—	—	1	40	.150
32	Podsakoff and Todor (1985)	—	0.56	—	1	827	.730
33	Putti (1985)	—	0.56	1	2	18	.490

continued

TABLE 1: continued

Number	*Study*	r_{xx}	r_{yy}	*TI*[a]	*Level*[b]	N[c]	*Effect*
34	Schriesheim (1980)	.90	0.84	—	1	308	.240
35	Shaw and Shaw (1962)	—	—	1	2	6	.230
36	Sheridan (1985)	.78	—	—	1	84	.300
37	Sheridan, Vredenburgh, and Abelson (1984)[f]	.75	0.78	—	1	327	.110
38	Sheridan, Vredenburgh, and Abelson (1984)[f]	.75	0.78	—	1	174	.200
39	Steel, Shane, and Kennedy (1990)	.80	0.97	—	1	69	−.040
40	Stinson and Hellebrandt (1972)[f]	—	—	1	2	11	.110
41	Stinson and Hellebrandt (1972)[f]	—	—	1	2	14	.000
42	Terborg, Castore, and DeNinno (1976)	—	—	1	2	42	−.038
43	Torrance (1955)	—	—	2	2	40	.262
44	Tziner and Vardi (1983)	—	—	2	2	115	.320
45	Widmeyer and Martens (1978)	—	1.00	2	2	66	.440
46	Williams and Hacker (1982)	—	1.00	2	2	9	.800
47	Williams and Widmeyer (1991)	—	—	1	2	18	.410
48	Wolfe and Box (1988)	.85	0.82	1	2	36	.320
49	Zaccaro (1991)	—	—	—	1	333	.260[d]
50	Zaccaro and Lowe (1988)	—	—	1	2	54	.443
51	Zaccaro and McCoy (1988)	—	—	1	2	132	.252

NOTE: r_{xx} = reliability of the measure for cohesion; r_{yy} = reliability of the measure for performance. Average reliabilities were 0.836 for r_{xx} and .832 for r_{yy}. These were the figures used to correct effect sizes and confidence intervals for all analyses.
a. 1 = low task interdependence; 2 = high task interdependence; M = task manipulated. Only group-level studies were coded for task interdependence because the use of an individual level of analysis for a group variable was not appropriate.
b. 1 = individual; 2 = group.
c. The *N* for individual-level studies represents the total number of individuals within groups. The *N* for group-level studies represents the total number of groups in the study.
d. Task and social cohesion were separately measured, and the two effect sizes were averaged together when conducting the meta-analyses, because a moderator analysis could not be conducted.
e. Studies 1, 2, and 3 respectively.
f. Groups 1 and 2 respectively.

(1978) and the self-report studies from the meta-analysis resulted in the modified-full analysis, with 44 effect sizes. Note that the full and modified-full analyses are composed of both individual- and group-level studies, as reflected in the sample sizes.

In the third step, the 44 remaining effect sizes relating cohesion and objective measures of performance were divided into two groups: those analyzing the cohesion-performance relationship at the group level (group analysis) and those analyzing the cohesion-performance relationship at the individual level (individual analysis).

In the final analysis for the effects of task interdependence, only studies that analyzed the effects of group cohesion at the group level of analysis were retained. This was done for two reasons. First, failure to separate the studies into different levels of analysis prior to conducting the analysis for task interdependence would confound the effects of level of analysis with the effects of task interdependence. Second, use of an individual level of analysis is inconsistent with theoretical conceptualizations of cohesion as a group-level construct. The 33 group-level effect sizes were divided into low and high task interdependence for the final moderator analysis.

RESULTS

The results of the meta-analyses can be found in Table 2. For the full meta-analysis, the effect size was .166 ($p \leq 6.6^{-42}$), and .199 after correction for error of measurement in both independent (average $r_{xx} = .836$) and dependent variables (average $r_{yy} = .832$). The lower and upper corrected endpoints of the confidence interval were .139 and .259, respectively. Sampling error alone accounted for only 12.1% of the variance in observed effect sizes, suggesting the existence of one or more moderators. A computation of the fail-safe *N* for the uncorrected effect size was conducted (Hunter & Schmidt, 1990; Orwin, 1983; Rosenthal, 1979). It was found that an additional 6,242 studies with a null effect size and an average sample size of 238 would have to be located to bring the obtained effect size to a just significant level of $p = .05$. It is highly unlikely that such a large body of studies could exist, and it lends confidence to the interpretation of the obtained findings.

For the modified-full analysis, the uncorrected weighted mean correlation was .221 ($p \leq 2.2^{-26}$), and the corrected correlation

TABLE 2: **Results of the Meta-Analyses**

Analysis	*Effects Included*[a]	*Total* N	k	$\bar{r}$	*Corr.* $\bar{r}$	*Corr.* SD $\bar{r}$	*Corr. Low 95%*	*Corr. High 95%*	*% Vr. Samp.*	χ^2	df
Full	1-51	12,115	51	.166	.199	.171	.139	.259	12.1	423	50
Modified-full	1-9, 11-16, 18-25, 28-31, 35-51	2,780	44	.221	.265	.091	.211	.318	63.9	69	43
Individual	4, 6, 9, 19, 20, 31,36-39, 49	1,634	11	.190	.228	.013	.176	.281	97.5	11	10
Group	1-3, 5, 7, 8, 11-16, 18, 21-25, 28-30, 35, 40-48, 50, 51	1,146	33	.264	.317	.131	.231	.401	59.9	55	32
Low task interdependence	1, 5, 7, 8, 11-13, 15, 21-23, 28, 30, 35, 40-42, 47, 48, 50, 51	683	21	.172	.206	.080	.109	.305	82.3	26	20
High task interdependence	2, 3, 14, 16, 24, 25, 29, 43-46	447	11	.387	.464	.028	.366	.562	95.8	12	10

NOTE: Total N = total sample size across studies; k = total number of studies included in analysis; $\bar{r}$ = weighted average effect size; Corr. $\bar{r}$ = weighted average effect size, corrected for unreliability; Corr. *SD* $\bar{r}$ = *standard deviation of r*, corrected for sampling error; Corr. Low 95% = lower endpoint of 95% confidence interval, corrected for unreliability; Corr. High 95% = upper endpoint of 95% confidence interval, corrected for unreliability; % Vr. Samp. = percentage of observed variance because of sampling error.

a. Effect numbers correspond to the numbering in first column of Table 1.

was .265. A computation of the fail-safe *N* indicated that an additional 2,164 studies with an average sample size of 63 and a null effect size would have to be found to bring the results to a just significant level of $p = .05$. It is highly unlikely that such a large body of unknown studies could exist, lending credibility to the obtained results. The lower and upper corrected endpoints of the confidence interval were .211 and .318, respectively. Sampling error alone accounted for 63.9% of the variance in observed effect sizes.

The difference in effect size and variance accounted for by sampling error illustrates how heavily the Hoiberg and Pugh (1978) and self-report studies affected the obtained results of the full analysis. Note also that the corrected mean effect sizes found in the full and modified-full analyses (.199 and .265, respectively) were substantially lower than the correlation obtained in Evans and Dion's (1991) meta-analysis (corrected correlation = .419, group level only), but they are comparable to Mullen and Copper's (1994) estimate of .248 using both individual- and group-level studies.

The group versus individual analysis provides some insight into the nature of the discrepancies between the results of the various meta-analyses. For the individual analysis, the uncorrected and corrected mean correlations were .190 and .228, respectively. The corrected 95% confidence interval ranged from .176 to .281, and sampling error accounted for 97.5% of the observed variance in the individual analysis effect sizes. For the group analysis, the uncorrected and corrected mean correlations were .264 and .317, respectively. The corrected 95% confidence interval ranged from .231 to .401, and sampling error alone accounted for 59.9% of the observed variance in the group effect sizes.

Note that the confidence intervals for the individual and group effect sizes have very little overlap. Additionally, the group effect size estimate is closer to Evans and Dion's (1991) finding, which was based only on group-level studies, and more discrepant from Mullen and Copper's (1994) finding, which was based on both individual- and group-level studies. Also, when compared to the modified-full analysis, the average standard deviation corrected for sampling error decreased (.0719 as compared to .0907). All of these findings provide strong support for the presence of a level-of-analysis moderator.

The last set of analyses investigated the relationship between cohesion and performance at low and high levels of task interdependence for group-level measures. The uncorrected and corrected effect sizes of cohesion at low levels of task interdependence were .172 and .206, respectively. The lower and upper ends of the corrected confidence interval were .109 and .305, respectively. Sampling error accounted for 82.3% of the variance observed. For high levels of task interdependence, the uncorrected and corrected effect sizes were .387 and .464, respectively. The lower and upper ends of the corrected confidence interval were .366 and .562, respectively, and sampling error accounted for 95.8% of the variance observed. Note that the confidence intervals for low and high levels of task interdependence do not overlap at all. Also, the average standard deviation for the task interdependence analysis was smaller than the group-level standard deviation (.0542 compared with .1311). These findings provide strong support for the existence of task interdependence as a moderator of the cohesion-performance relationship, in contrast to the finding obtained by Mullen and Copper (1994) for interaction requirement in which level of analysis was not considered.

DISCUSSION

SUMMARY OF FINDINGS

Although it appears that cohesion and performance are generally positively related, the results of this study suggest that a more complex relationship exists. Specifically, a moderator analysis for task interdependence found that the magnitude of the cohesion-performance relationship varied considerably across tasks characterized as either high or low on task interdependence. In addition, the results suggest that the broad range of obtained findings in the literature is due in part to variability in the level of analysis used to conceptualize and operationalize the cohesion construct.

The finding that level of analysis accounts for variation in effect sizes highlights the fundamental need to acknowledge levels-of-analysis issues inherent in research on groups. Work from the levels perspective has identified the need to match the level of statistical analysis with the level of theory (i.e., level of generalization) to draw correct conclusions. In other words, if one desires to talk about groups, one needs to use group-level measures in the analysis or use appropriate aggregation procedures for individual-level measurements (Rousseau, 1985). In our review of the literature, we found 11 studies that (explicitly or implicitly) viewed cohesion as something other than a group-level construct and operationalized its measurement at the individual level. However, many of those same studies then went on to discuss findings as though the group was the focal unit of analysis. These procedures are likely to lead to fallacious conclusions about the constructs of interest (Roberts et al., 1978; Rousseau, 1985). George and Bettenhausen's (1990) study is a noteworthy exception; they were attentive to levels issues in their study and properly used aggregation procedures to create group constructs.

Another issue that is highlighted by the levels framework adopted in this study is the casual aggregation of individual-level data to group-level means. Although a number of studies used group means as group-level measures, most failed to examine agreement at the individual level before aggregating. If individual-level data are summarized as group means without ensuring the homogeneity of responses at the individual level, then aggregation bias becomes a potentially severe problem. Only one study was found that examined agreement before aggregating individual responses to questionnaire items to group means (George & Bettenhausen, 1990). Because of the widespread failure to examine agreement before aggregation, we were unable to separate the effects of aggregation bias from the true nature of the cohesion-performance relationship.

The results of the task interdependence analysis suggest that the strength of the cohesion-performance relationship is substantially determined by the nature of the task. When the demands of the task necessitate coordination, communication, and mutual performance monitoring among group members, cohesion and performance are more strongly related than when task interdependence is low.

It is interesting to note that Mullen and Copper (1994) did not find an effect for interaction requirement in their analyses. There are two possible causes of the discrepancy between the current findings for task interdependence and Mullen and Copper's interaction requirement. First, in the current study, the effect of level of analysis was removed by only including group-level studies in our analysis of task interdependence. As noted, this could substantially change the nature of obtained findings. Second, it is unclear what coding scheme was used by Mullen and Copper to categorize tasks into low and high group interaction requirement. We used constructs derived from previous research (e.g., communication, coordination, and mutual performance monitoring) to perform our coding analysis. It should be noted, however, that results from our analysis of task interdependence are consistent with previous findings for task type (e.g., sports, military, etc.) obtained by Mullen and Copper and by Gully et al. (1993). Although several existing models address the cohesion-performance relationship in general (e.g., Carron, 1982; Carron, Widmeyer, & Brawley, 1985), our results suggest that task interdependence should be included in future models.

FUTURE RESEARCH DIRECTIONS

Although the current study has obtained interesting findings, there are factors concerning these analyses that warrant caution. For example, the sparse task descriptions provided by most studies in the literature necessitated coding task interdependence in a global fashion (i.e., high vs. low). Future studies involving cohesion and performance should provide richer descriptions of the task setting to allow future researchers to examine subsets of the theoretically meaningful variables in group-level contexts.

It should also be noted that this study has not attempted (nor is it able) to identify the direction of the presumed causal relationship between cohesion and performance. Despite the attempt by Mullen and Copper (1994) to do so, the current number of studies within a given time period is too small to allow such an analysis. Future research should examine the possibility of a dynamic, reciprocal relationship between the two constructs through the use of longitudinal designs.

Another aspect of the cohesion-performance relationship that should be addressed is the multidimensional nature of cohesion and its differential effects on performance. The task/social distinction has been made by a number of researchers and has received theoretical and empirical support (e.g., Carron, 1988; Johnson & Fortman, 1988; Mullen & Copper, 1994; Widmeyer, Brawley, & Carron, 1985, 1990; Zaccaro & Lowe, 1988; Zaccaro & McCoy, 1988). Although both social and task cohesion may be important and independent determinants of performance, Mullen and Copper (1994) suggested task cohesion is the stronger determinant of the two. However, it should be noted that there are only a small number of empirical studies (e.g., Widmeyer et al., 1990) that have explicitly examined how these two types of cohesion affect group processes and performance, particularly for different types of tasks.

Future research should also investigate the effects of group goals or norms on the relationship between cohesion and performance. If groups are highly cohesive and group goals are congruent with organizational goals, then the effectiveness of the group as measured by organizational standards should be very high. When the group is highly cohesive and group goals are not congruent with organizational goals, then performance is likely to be very low (Greene, 1989; Seashore, 1954; Stogdill, 1972). Researchers have discussed group goals or norms as an important moderator variable, but little research investigated its effect empirically.

In summary, we feel that this study makes two important contributions. First, variance in obtained cohesion-performance effect sizes can be partially explained by the manner in which cohesion was conceptualized and treated during analysis. Future research should conceptualize cohesion as a group-level construct and treat it accordingly, attending to aggregation issues when necessary. Second, task interdependence was found to strongly affect the strength of the cohesion-performance relationship. The cohesion-performance relationship was found to be stronger for tasks requiring high interdependence than tasks involving low interdependence.

As these issues are explored, the relationship between cohesion and group processes can be developed. By clarifying the definition, operationalization, and analysis of group cohesion through the adoption of a levels perspective, and taking into account the constraining aspects of the task performed, we hope to facilitate research that will increase our understanding of the construct and lead to useful practical applications in service, transportation, military, sports, and other such contexts.

REFERENCES

Bakeman, R., & Helmreich, R. (1975). Cohesiveness and performance: Covariation and causality in an undersea environment. *Journal of Experimental Social Psychology, 11*, 478–489.

Ball, J. R., & Carron, A. V. (1976). The influence of team cohesion and participation motivation upon performance success in intercollegiate ice hockey. *Canadian Journal of Applied Sport Sciences, 1*, 271–275.

Bettenhausen, K. (1991). Five years of groups research: What we have learned and what needs to be addressed. *Journal of Management, 17*, 345-381.

Bird, A. M. (1977). Team structure and success as related to cohesiveness and leadership. *Journal of Social Psychology, 103*, 217–223.

Carron, A. (1982). Cohesiveness in sport groups: Interpretations and considerations. *Journal of Sport Psychology, 4*, 123–138.

Carron, A. (1988). *Group dynamics in sports*. Ontario, Canada: Spodym.

Carron, A., Widmeyer, W., & Brawley, L. (1985). The development of an instrument to assess cohesion in sport teams: The Group Environment Questionnaire. *Journal of Sport Psychology, 7*, 244–266.

Cartwright, D. (1968). The nature of group cohesiveness. In D. Cartwright & A. Zander, (Eds.), *Group dynamics: Research and theory* (3rd ed.). New York: Harper & Row.

Cattell, R. (1948). Concepts and methods in the measurement of group syntality. *Psychological Review, 55*, 48–63.

Cohen, D., Whitmyre, J. W., & Funk, W. H. (1960). Effect of group cohesiveness and training upon creative thinking. *Journal of Applied Psychology, 44*, 319–322.

Colarelli, S. M., & Boos, A. L. (1992). Sociometric and ability-based assignment to work groups: Some implications for personnel selection. *Journal of Organizational Behavior, 13*, 187–196.

Dailey, R. C. (1978). Perceived group variables as moderators of the task characteristics-individual performance relationship. *Journal of Management, 4*, 69–80.

Darley, J. G., Gross, N., & Martin, W. C. (1952). Studies of group behavior: Factors associated with the predictability of groups. *Journal of Applied Psychology, 36*, 396–403.

Davis, J. (1969). *Group performance*. Reading, MA: Addison-Wesley.

Deep, S., Bass, B., & Vaughan, J. (1967). Some effects on business gaming of previous quasi-T group affiliations. *Journal of Applied Psychology, 51*, 426–431.

Dorfman, P. W., & Stephan, W. G. (1984). The effects of group performance on cognitions, satisfaction, and behavior: A process model. *Journal of Management, 10*, 173–192.

Evans, C. R., & Dion, K. L. (1991). Group cohesion and performance: A meta-analysis. *Small Group Research, 22*, 175–186.

Fandt, P. M. (1991). The relationship of accountability and interdependent behavior to enhancing team consequences. *Group and Organization Studies, 16*, 300–312.

Fiedler, F. E. (1954). Assumed similarity measures as predictors of team effectiveness. *Journal of Abnormal and Social Psychology, 49*, 381–388.

Gekoski, N. (1952). Predicting group productivity. *Personnel Psychology, 5*, 281–292.

George, J. M., & Bettenhausen, K. (1990). Understanding prosocial behavior, sales performance, and turnover: A group-level analysis in a service context. *Journal of Applied Psychology, 75*, 698–709.

Goodacre, D. M. (1951). The use of a sociometric test as a predictor of combat unit effectiveness. *Sociometry, 14*, 148–152.

Greene, C. (1989). Cohesion and productivity in work groups. *Small Group Behavior, 20*, 70–86.

Gully, S. M., Whitney, D. J., & Devine, D. J. (1993, May). *Bringing cohesion to group cohesion research: A literature review and meta-analysis*. Paper presented at the eighth annual conference of the Society for Industrial and Organizational Psychology, San Francisco.

Hemphill, J. K., & Sechrest, L. B. (1952). A comparison of three criteria of aircrew effectiveness in combat over Korea. *Journal of Applied Psychology, 36*, 323–327.

Hoiberg, A., & Pugh, W. (1978). Predicting Navy effectiveness: Expectations, motivation, personality, aptitude, and background variables. *Personnel Psychology, 31*, 841–852.

Hoogstraten, J., & Voorst, H. (1978). Group cohesion, task performance, and the experimenter expectancy effect. *Human Relations, 31*, 939–956.

Hunter, J., & Schmidt, F. (1990). *Methods of meta-analysis*. Newbury Park, CA: Sage.

Indik, B. (1968). The scope of the problem and some suggestions toward a solution. In B. P. Indik & F. K. Berrien (Eds.), *People, groups and organizations*. New York: Teachers College Press.

Jaffe, E., & Nebenzahl, I. (1990). Group interaction and business game performance. *Simulation & Gaming, 21*, 133–146.

James, L. (1982). Aggregation bias in estimates of perceptual agreement. *Journal of Applied Psychology, 67*, 219–229.

James, L. R., Demaree, R. G., & Wolf. G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology, 69*, 85–98.

Johnson, M., & Fortman, J. (1988). Internal structure of the Gross Cohesiveness Scale. *Small Group Behavior, 19*, 146–152.

Keller, R. T. (1986). Predictors of the performance of project groups in R and D organizations. *Academy of Management Journal, 29*, 715–726.

Keyton, J., & Springston, J. (1990). Redefining cohesiveness in groups. *Small Group Research, 21*, 234–254.

Klein, K., & Christiansen, G. (1969). Group composition, group structure, and group effectiveness of basketball teams. In J. W. Loy & G. S. Kenyon (Eds.), *Sport, culture, and society: A reader on the sociology of sport* (pp. 397–408). London: Macmillan.

Klein, K. J., Dansereau, F., & Hall, R. J. (1994). Levels issues in theory development, data collection, and analysis. *Academy of Management Journal, 19,* 195–229.

Kozlowski, S., & Hattrup, K. (1992). A disagreement about within-group agreement: Disentangling issues of consistency versus consensus. *Journal of Applied Psychology, 77,* 161–167.

Landers, D. M., Wilkinson, M. O., Hatfield, B. D., & Barber, H. (1982). Causality and the cohesion-performance relationship. *Journal of Sport Psychology, 4,* 170–183.

Levine, J. M., & Moreland, R. L. (1990). Progress in small group research. *Annual Review of Psychology, 41,* 585–634.

Littlepage, G., Cowart, L. & Kerr, B. (1989). Relationships between group environment scales and group performance and cohesion. *Small Group Behavior, 20,* 50–61.

Lott, A., & Lott, B. (1965). Group cohesiveness as interpersonal attraction: A review of relationships with antecedent and consequent variables. *Psychological Bulletin, 64,* 259–309.

Melnick, M. J., & Chemers, M. M. (1974). Effects of group structure on the success of basketball teams. *Research Quarterly, 45,* 1–8.

Mitchell, T. R., & Silver, W. S. (1990). Individual and group goals when workers are interdependent: Effects on task strategies and performance. *Journal of Applied Psychology, 75,* 185–193.

Moos, R. & Speisman, J. (1962). Group compatibility and productivity. *Journal of Abnormal and Social Psychology, 65,* 190–196.

Mossholder, K. W., & Bedeian, A. G. (1983). Cross-level inference and organizational research: Perspectives on interpretation and application. *Academy of Management Review, 8,* 547–558.

Mudrack, P. (1989a). Defining group cohesiveness: A legacy of confusion? *Small Group Behavior, 20,* 37–49.

Mudrack, P. (1989b). Group cohesiveness and productivity: A closer look. *Human Relations, 42,* 771–785.

Mullen, B., & Copper, C. (1994). The relation between group cohesiveness and performance: An integration. *Psychological Bulletin, 115,* 210–227.

Norris, D., & Niebuhr, R. (1980). Group variables and gaming success. *Simulation and Games, 11,* 301–312.

Orwin, R. H. (1983). A fail-safe *N* for effect size in meta-analysis. *Journal of Educational Psychology, 8,* 157–159.

Ostroff, C. (1993). Comparing correlations based on individual-level and aggregated data. *Journal of Applied Psychology, 78,* 569–582.

Piper, W., Marrache, M., Lacroix, R., Richardsen, A., & Jones, B. (1983). Cohesion as a basic bond in groups. *Human Relations, 32,* 93–108.

Podsakoff, P. M., & Todor, W. D. (1985). Relationships between leader reward and punishment behavior and group processes and productivity. *Journal of Management, 11,* 55–73.

Putti, J. M. (1985). Leader behavior and group characteristics in work improvement teams—The Asian context. *Public Personnel Management, 14,* 301–306.

Roberts, K. H., Hulin, C. L., & Rousseau, D. M. (1978). *Developing an interdisciplinary science of organizations.* San Francisco: Jossey-Bass.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review, 15,* 351–357.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin, 86,* 638–641.

Rousseau, D. M. (1985). Issues of level in organizational research: Multi-level and cross-level perspectives. In L. L. Cummings & B. M. Staw (Eds.), *Research in organizational behavior* (Vol. 7, pp. 1–37), Greenwich, CT: JAI.

Saavedra, R., Earley, P., & Van Dyne, L. (1993). Complex interdependence in task-performing groups. *Journal of Applied Psychology, 78*, 61–72.

Salas, E., Dickinson, T. L., Converse, S. A., & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. In R. W. Swezey & E. Salas (Eds.), *Teams: Their training and performance* (pp. 3–29). Norwood, NJ: Ablex.

Schriesheim, J. (1980). The social context of leader-subordinate relations: An investigation of the effects of group cohesiveness. *Journal of Applied Psychology, 65*, 183–194.

Seashore, S. E. (1954). *Group cohesiveness in the industrial work group*. Ann Arbor: Institute for Social Research, University of Michigan.

Shaw, M. E. (1971). *Group dynamics*. New York: McGraw-Hill.

Shaw, M., & Shaw, L. (1962). Some effects of sociometric grouping upon learning in a second-grade classroom. *Journal of Social Psychology, 57*, 453–458.

Shea, G. P., & Guzzo, R. A. (1987). Groups as human resources. In K. M. Rowland & G. R. Ferris (Eds.), *Research in personnel and human resources management* (Vol. 5, pp. 323–356). Greenwich, CT: JAI.

Sheridan, J. E. (1985). A catastrophe model of employee withdrawal leading to low job performance, high absenteeism, and job turnover during the first year of employment. *Academy of Management Journal, 28*, 88–109.

Sheridan, J. Vredenburgh, D. J., & Abelson, M. (1984). Contextual model of leadership influence in hospital units. *Academy of Management Journal, 27*, 57–78.

Steel, R. P. Shane, G. S., & Kennedy, K. A. (1990). Effects of social-system factors on absenteeism, turnover, and job performance. *Journal of Business and Psychology, 4*, 423–430.

Stinson, J. E., & Hellebrandt, E. T. (1972). Group cohesiveness, productivity, and strength of formal leadership. *Journal of Social Psychology, 87*, 99–105.

Stogdill, R. (1972). Group productivity, drive, and cohesiveness. *Organizational Behavior and Human Performance, 8*, 26–43.

Sundstrom, E., De Meuse, K. P., & Futrell, D. (1990). Work teams: Applications and effectiveness. *American Psychologist, 45*, 120–133.

Terborg, J. R., Castore, C., & DeNinno, J. A. (1976). A longitudinal field investigation of the impact of group composition on group performance and group cohesion. *Journal of Personality and Social Psychology, 34*, 782–790.

Thorndike, E. L. (1939). On the fallacy of imputing the correlations found for groups to the individuals or smaller groups composing them. *American Journal of Psychology, 52*, 122–124.

Tinsley, H.E.A., & Weiss, D. J. (1975). Interrater reliability and agreement of subjective judgments. *Journal of Counseling Psychology, 22*, 358–376.

Torrance, E. P. (1955). Perception of group functioning as a predictor of group performance. *Journal of Social Psychology, 42*, 271–282.

Tziner, A., & Vardi, Y. (1983). Ability as a moderator between cohesiveness and tank crews performance. *Journal of Occupational Behavior, 4* 137–143.

Wanous, J. P., Sullivan, S. E., & Malinak, J. (1989). The role of judgment calls in meta-analysis. *Journal of Applied Psychology, 74*, 259–264.

Widmeyer, N., Brawley, L., & Carron, A. (1985). *The measurement of cohesion in sport teams: The Group Environment Questionnaire*. London, Ontario, Canada: Sports Dynamics.

Widmeyer, N., Brawley, L., & Carron, A. (1990). The effects of group size in sport. *Journal of Sport and Exercise Psychology, 12*, 177–190.

Widmeyer, N., Brawley, L., & Carron, A. (1992). Group dynamics in sport. In T. S. Horn (Ed.), *Advances in sport psychology* (pp. 124–139). Champaign, IL: Human Kinetics.

Widmeyer, W. N., & Martens, R. (1978). When cohesion predicts performance outcome in sport. *Research Quarterly, 49*, 372–380.

Williams, J. M., & Hacker, C. M. (1982). Causal relationships among cohesion, satisfaction, and performance in women's intercollegiate field hockey teams. *Journal of Sport Psychology, 4*, 324–337.

Williams, J., & Widmeyer, W. (1991). The cohesion-performance outcome relationship in a coaching sport. *Journal of Sport and Exercise Psychology, 13*, 364–371.

Wolfe, J., & Box, T. (1988). Team cohesion effects on business game performance. *Simulations and Games, 19*, 82–98.

Zaccaro, S. (1991). Non-equivalent associations between forms of cohesiveness and group-related outcomes: Evidence for multidimensionality. *Journal of Social Psychology, 131*, 387–399.

Zaccaro, S., & Lowe, C. (1988). Cohesiveness and performance on an additive task: Evidence for multidimensionality. *Journal of Social Psychology, 128*, 547–558.

Zaccaro, S., & McCoy, M. C. (1988). The effects of task and interpersonal cohesiveness on performance of a disjunctive group task. *Journal of Applied Social Psychology, 18*, 837–851.

Zander, A. (1979). The psychology of group processes. *Annual Review of Psychology, 30*, 417–451.

Stanley M. Gully is a Ph.D. candidate in the Industrial/Organizational Psychology Program at Michigan State University. His current interests include the effects of leadership on group/team processes and performance, and team training and decision-making effectiveness.

Dennis J. Devine is a Ph.D. candidate in the Industrial/Organizational Psychology Program at Michigan State University. His interests include understanding expert-novice differences, factors affecting group/team performance in complex environments, and individual selection in team contexts.

David J. Whitney is a Ph.D. candidate in the Industrial/Organizational Psychology Program at Michigan State University. His current research examines item bias and factors influencing the effectiveness of volunteer, real-world teams.