The Effectiveness of Managerial Leadership Development Programs: A Meta-Analysis of Studies from 1982 to 2001

Doris B. Collins, Elwood F. Holton III

Eighty-three studies from 1982 to 2001 with formal training interventions were integrated via meta-analytic techniques to determine the effectiveness of interventions in their enhancement of performance, knowledge, and expertise at the individual, team or group, or organizational level. The studies were separated by research design, with the outcome measure of the intervention as the unit of analysis. The effect size for knowledge outcomes ranged from .96 to 1.37; expertise outcomes from .35 to 1.01; and system outcomes averaged .39. Interventions with knowledge outcomes were found to be more effective than in the Burke and Day (1986) meta-analysis, with the most effective interventions using a single group pretest-posttest research design. Methodological and conceptual differences in Burke and Day's meta-analysis on the effectiveness of managerial training make historical comparisons risky. The data suggest that practitioners can attain substantial improvements in both knowledge and skills if sufficient front-end analysis is conducted to assure that the right development is offered to the right leaders.

Many organizations are concerned about the leadership inadequacies of their employees and are committed to education and training to develop managers' skills, perspectives, and competencies (Conger & Benjamin, 1999). Leadership development literature indicates that significant financial payoffs are found among companies that emphasize training and development (Huselid, 1995; Jacobs & Jones, 1995; Lam & White, 1998; Swanson, 1994; Ulrich, 1997). Organizations are now realizing that workplace expertise is crucial to maintaining optimal performance and adapting to change in today's dynamic business world (Herling, 2000; Krohn, 2000). As companies "recognize the shortage of talented managers and the importance of developing 'bench strength' to

Human Resource Development Quarterly, vol. 15, no. 2, Summer 2004

widen perspectives to compete globally" (p. xii), budgets for leadership development programs are expected to grow (Gibler, Carter, & Goldsmith, 2000). The interest in this research was to determine whether the effectiveness of leadership development programs continues to lag behind the demand curve for leaders, as Klenke (1993) believed. Also, Lynham (2000) indicated that the field of leadership development could benefit from further purposeful and scholarly inquiry and study. In addition, Lynham stated that there was a "need to gather up studies and understand leadership development and to conduct analyses of the evolution and nature or what is really known in this field" (p. 5). The authors recognized that a meta-analysis could serve as a useful statistical approach for making sense out of leadership development studies and to determine the effectiveness of interventions.

Even though leadership development interventions are pervasive, research also indicates that organizations are spending little time evaluating the effectiveness of their interventions and, more specifically, evaluating whether those programs improve the organization's performance (Sogunro, 1997). That leadership development efforts will result in improved leadership skills appears to be taken for granted by many corporations, professional management associations, and consultants. In essence, many companies naively assume that leadership development efforts improve organizational efforts. *Leadership development* is defined as "every form of growth or stage of development in the life cycle that promotes, encourages, and assists the expansion of knowledge and expertise required to optimize one's leadership potential and performance" (Brungardt, 1996, p. 83).

There are many opinions as to why organizations are not evaluating or reporting the results of their leadership development interventions. First, the competencies required to be an accomplished leader are complex and overlapping (Collins, Lowe, & Arnett, 2000). Second, McCauley, Moxley, and Van Velsor (1998) suggested that a full range of leadership development experiences includes mentoring, job assignments, feedback systems, on-the-job experiences, developmental relationships, exposure to senior executives, leader-follower relationships, and formal training. While the variety of tasks and challenges encountered on the job are a major source of learning, the reality is that all jobs are not developmentally equal (McCauley & Brutus, 1998), nor can they be expressed in an objective manner, which makes evaluation more difficult. Third, organizations appear to believe that improving knowledge and skills of individual employees automatically enhances the organization's effectiveness. What are measured most are the interpersonal skills and the work performance of individual managers (Moxnes & Eilertsen, 1991). Measurement of organizational effectiveness is somewhat more difficult, because it often involves analysis at multiple levels of the organization (Rummler & Brache, 1995). Fourth, some researchers believe that evaluative studies of leadership development are sparse because of the lack of an evaluation model that adequately measures the effect of the interventions on the performance of the

organization (Alliger & Janak, 1989; Bassi, Benson, & Cheney, 1996; Clement, 1982; Holton, 1996; Moller & Mallin, 1996; Newstrom, 1995; Swanson, 1998). Kirkpatrick's model is simple, and has been used primarily to evaluate reactions, learning, and behavior, all of which are measurement of transfer of training to individual employees (Alliger & Janak, 1989). However, Kirkpatrick's model does not appear to be effective in measuring organizational performance, the effectiveness of an organization in achieving outcomes as identified by its strategic goals, or the realization of a return on investments (Holton, 1999).

The meta-analysis conducted by Burke and Day (1986) is commonly regarded as the principal empirical support for the effectiveness of managerial training and leadership development programs. Burke and Day's meta-analysis included seventy published and unpublished studies from business and industry spanning thirty years (1951–1982). Studies involved managerial or supervisory personnel, evaluated the effectiveness of more than one training program, and included at least one control or comparison group. Burke and Day found that managerial training was moderately effective and provided true mean effect sizes (in parentheses) for each of the four criterion-measure categories used: subjective learning (.34), objective learning (.38), subjective behavior (.49), and objective results (.67). Burke and Day's study clarified the breadth of managerial training, but indicated that more empirical research was needed before conclusive statements could be made. They found that managerial training was pervasive and primarily focused on improving individual managerial skills and on-the-job performance. The lack of evaluative research caused Burke and Day to believe that organizations were unaware of the effectiveness of management training programs in improving job performance. Other significant conclusions from the Burke and Day study were (1) researchers need to improve reports that evaluate organizational interventions to provide cumulative analyses of the effectiveness of managerial training; (2) trainers and organizational decision makers should not rely on training program content area descriptions when choosing the utility of managerial training programs; (3) the level of experience of the trainer may be significant in influencing the effectiveness of the training program; and (4) different management training methods do not necessarily lead to increased knowledge and improved performance. They also found that short time frames and reliance on self-report measures typified management development research. It is important to note that only two of their studies used organizational variables as outcome criteria.

Zhang's (1999) unpublished meta-analysis of forty-seven studies most closely replicated Burke and Day's meta-analysis and found that management training made a difference, but also called for further research. Several other meta-analyses on topics related to leadership development have been published (Bayley, 1988; Chen, 1994; Lai, 1996; Leddick, 1987). Bayley (1988) reported highly significant effects of continuing education on behavioral

change in clinical practices. Chen (1994) statistically integrated studies regarding the effectiveness of cross-cultural interventions to be effective. Lai (1996) integrated findings of twelve studies on the program effectiveness of educational leadership training, using only experimental or quasi-experimental design, and found that educational leadership training had a small effect when leader behavior changes were measured. Leddick (1987) found that knowledge objectives seemed to be associated with stronger productivity improvements than other types of objectives. Across all studies and multiple research design types in Leddick's meta-analysis, the analysis produced an overall effect size of .67, with a .98 effect size for managers only. Chen and Leddick also discovered that control group studies produced lower effect sizes than single group pretest-posttest studies.

There have been enough changes in leadership and leadership development since Burke and Day's meta-analysis (1986) that an additional meta-analysis is warranted to understand the effectiveness of current leadership development programs. Leadership development is "no longer focused on the individual learner but increasingly on shaping the worldviews and behaviors of cohorts of managers and, even transforming entire organizations" (Conger & Benjamin, 1999, p. xii). Strategic vision is now a focus of leaders because of the almost continuous restructuring activities, demographic changes in the workforce, and technological changes in a more complex and fast-paced system (Friedman, 2000; Gibler, Carter, & Goldsmith, 2000; Hooijberg, Hunt, & Dodge, 1997). The ability of multinational companies to compete in the global market is contingent upon their ability to change and adapt resources strategically (Caligiuri & Stroh, 1995). Global organizations are also faced with dual reporting structures, proliferation of communication channels, overlapping responsibilities, and barriers of distance, language, time, and culture (Friedman, 2000), but Marquardt and Engel (1993) found that very few leadership development interventions have a global focus. Organizations today face a multitude of competing, outcome-based demands (Levi & Mainstone, 1992)—ones that stem not only from customer demands but from a variety of forces, such as federal mandates and national accreditation standards. Peter Vaill (1990) used the metaphor of "permanent white water" to represent the uncertainty, chaos, and complexity inherent in today's managerial environment.

Since Burke and Day's research, transformational and team leadership (Bass, 1985; Hackman & Walton, 1986; Larson & LaFasto, 1989), 360-degree feedback (Lepsinger & Lucia, 1997), and on-the-job experiences (McCauley & Brutus, 1998) have been introduced into leadership development literature. Leadership development has undergone a shift in learning approaches and program design, and greater emphasis has been placed on groups of managers (Conger & Benjamin, 1999).

This meta-analysis adopted the term "managerial leadership development" to integrate the traditional managerial and leadership behaviors (Bass, 1990; Fleishman, Mumford, Zaccaro, Levin, Korotkin, & Hein, 1991; House & Aditya, 1997; Kotter, 1990; Yukl, 1989; Yukl & van Fleet, 1992) when those

behaviors are different but complementary. It also adopted the full range of leadership model (Avolio, 1999; Bass, 1998) in which all leaders' and managers' behaviors are different, but all leaders displayed both types of behavior to varying degrees, and transformational leadership augments transactional leadership. This research also subscribed to Holton and Naquin's (2000) definition of *high-performance leadership* as "leading and managing people and organizational systems to achieve and sustain high levels of effectiveness by optimizing goals, design and management at the individual, process and organizational levels" (p. 1) and used their competency model to define intervention content areas.

Because little is known about what knowledge and skills or processes in managerial leadership development interventions contribute to organizational performance (Campbell, Dunnette, Lawler, & Weick, 1970; Fiedler, 1996; Lynham, 2000), this research focused on outcomes in terms of knowledge, expertise, or system results at the individual, team or group, or organizational level (Rummler & Brache, 1995), with outcomes defined as "a measurement of effectiveness or efficiency (of the organization) relative to core outputs of the system, subsystem, process, or individual" (Holton, 1999, p. 33). Managerial leadership development outcomes have traditionally focused on individual learning and skills without regard to organizational performance, using Kirkpatrick's evaluation model (1998). However, "the missing elements and relationships in the Kirkpatrick model prohibit making accurate statements about system states" (Holton, 1996, p. 7). Therefore, the Results Assessment System (Swanson & Holton, 1999) was used to analyze the outcomes of leadership development studies from both a learning and performance (system) perspective. By integrating the results of leadership and management development research via meta-analytic techniques (Hunter & Schmidt, 1990; Lipsey & Wilson, 2001), this meta-analysis will assist future researchers in determining the effectiveness of managerial leadership development interventions and in their enhancement of organizational performance, individual knowledge, and expertise.

This study answered the following research questions: (1) Across studies measuring knowledge outcomes, how effective is managerial leadership development? (2) Across studies measuring expertise (or behavior) outcomes, how effective is managerial leadership development? (3) Across studies measuring system outcomes, how effective is managerial leadership development? (4) What moderator effects can be detected for the following variables: training content, organization type, job classification level, publication type, measurement method, research design, and objective-subjective outcomes?

Method

This section discusses meta-analytic methods used in this study. However, space limitations prevent an in-depth discussion of meta-analysis. Readers who want a more comprehensive knowledge of meta-analysis should consult Aguinis and

Pierce (1998); Hunter and Schmidt (1990); Lipsey and Wilson (2001); or Carlson and Schmidt (1999).

Literature Search. The literature search involved three steps: computerized search of all available databases, manual search of existing literature, and communication with subject matter experts to locate unpublished studies. Studies were located by conducting computer searches using WebSPIRS and Ingenta (UNCOVER) to search ERIC, PsychInfo, and Dissertation Abstracts International databases with keywords *effectiveness, impact, influence, outcomes,* and *results* that intersected with the key subject areas of *executive development, executive training, leadership development, leadership education, leadership training, management development, management education, management skills, management training, managerial training, supervisory training, supervisory development, 360-degree feedback, multisource feedback, multi-rater feedback, mentoring, coaching, and dyadic relationships.* In addition, a computer search was conducted of five Web sites identified by subject matter experts as ones likely to provide leadership development studies:

http://cls.binghamton.edu/library.htm
http://www.ari.army.mil
http://management.bu.edu/research/edrt/index.asp
www.grcl.com
leadership.center@boeing.com

A manual search was conducted of reference lists of all studies located through the computerized search; an article-by-article search was conducted of all volumes of *Journal of Applied Psychology, Academy of Management Journal, Personnel Psychology, Group and Organization Studies/Group and Organization Management, Organizational Behavior and Human Decision Processes/Organizational Behavior and Human Performance, Human Relations,* and *Journal of Vocational Behavior* from 1982 to 2001. The tables of contents were reviewed for all volumes of the following journals from 1982 to 2001: *Leadership Quarterly, Journal of Leadership Studies, Journal of Management Development, Organizational Dynamics, Human Resource Development Quarterly,* and *Human Resource Management.* All studies cited in *The Impact of Leadership* by Clark, Clark, and Campbell (1992) were reviewed.

A meta-analysis is not considered complete if a subset of the population is intentionally omitted. Efforts were taken to prevent the "file-drawer problem" where "journals are filled with five percent of the studies that show Type I error, while the file drawers are filled with 95 percent of the studies that show non-significant results" (Rosenthal, 1984, p. 107). A search for unpublished manuscripts was conducted to help ensure that findings from this meta-analysis were not biased due to the absence of unobserved and unobservable effect sizes (Lipsey & Wilson, 2001). Unpublished studies were sought through e-mail contacts with all known authors of leadership development

studies and individuals at the Center for Creative Leadership who were likely to have knowledge about available managerial leadership development studies. Presenters on leadership or management development at conferences of the Academy of Human Resource Development from 1998 to 2001, and the Society of Industrial and Organizational Psychology in 2000 were contacted. Because of the limited information within some studies, the degree of methodological rigor or quality of research design was not a selection criterion.

To be included, each study had to meet the following criteria:

1. The study was operationally defined as an organizational managerial leadership development study.
2. The study incorporated an *intervention* that involved managers, leaders, executives, officers, supervisors, and/or foremen, defined as a deliberately planned effort by an individual, group, or organization with the specific intent to enhance managerial leadership potential at the individual, group or team, or organizational level.
3. The study reported quantitative analyses from one of four research designs: posttest only control group (POWC); pretest-posttest with control group (PPWC); single group pretest-posttest (SGPP); or correlational (CORR).
4. The study described the treatment and outcome measures.
5. The study reported the group means and standard deviations, Cohen's *d*, probability level, *t*-value, Pearson's *r,* or raw data from which an effect size was determined, or the author provided this information when contacted.
6. The study was published in English between January 1982 and December 2001, and did not duplicate any studies that were used in Burke and Day's (1986) meta-analysis, as the purpose of the research was to pick up where they left off and analyze studies from that point in time forward. The current research attempted to locate all available leadership development studies. Because the researchers were not attempting to locate studies in every language, the search was of journals primarily written in English. Nevertheless, many of those journals have a high percentage of international authors, resulting in 28 percent of the interventions being from non–U.S. companies.

Overall, the literature searches located 346 articles that contained more than one keyword. However, 214 were not empirical studies and did not meet the criteria for inclusion in the study for one of the following reasons: (1) it described some theoretical aspect of management development, (2) defined training methods of an intervention, (3) summarized the developmental aspects of management positions, (4) described a naturally occurring process, (5) defined the behavioral change of a student group who participated in an intervention, or (6) described an intervention of non-managerial-level employees. The remaining

132 empirical studies were reviewed fully; twenty-nine of the studies were discarded because we were either unable to obtain adequate statistical information or they were duplicates of previous studies. Of the 103 studies remaining that met the criteria for inclusion, the low number of studies with feedback, developmental relationships, and on-the-job interventions prevented us from performing a statistically sound meta-analysis on studies with those intervention types. Thus, this meta-analysis was reduced to eighty-three studies with formal training interventions only.

Coding of Studies. A coding form captured the author's name, publication type and year, job classification level, organization type, country, program name, sample size, intervention type, content focus, outcome category, outcome variables measured, measurement instrument and method, research design, and statistical data. Effect sizes were calculated from means and standard deviation, Cohen's *d,* Pearson's *r, F* or *t*-value, or *p* levels as documented on the coding form. Detailed coding instructions were developed to train coders, ensuring consistency in the coding of studies.

Coding Verification. Initially, the primary researcher coded a random sample of twenty studies twice, with a 95 percent coding consistency. To determine accuracy of the researcher's coding, the researcher required two additional individuals with significant knowledge of leadership development to independently code the same sample, with 88 percent and 92 percent agreement with the researcher. Discrepancies were resolved based upon discussion, with 100 percent agreement. The remaining sixty-three studies were coded twice by the primary researcher to ensure accuracy.

Coding Characteristics. The four key study characteristics were intervention type, content focus, outcome category, and research design. The intervention type was defined by using a full range of managerial leadership development interventions (McCauley, Moxley, & Van Velsor, 1998) and were categorized as formal training; feedback; developmental relationships; and on-the-job experiences. This research focused on formal training programs, defined by the researcher as structured training programs in a formal setting either inside or outside the organization designed to develop the individual employee.

The high-performance leadership competency model (Holton & Naquin, 2000) was used to define intervention content areas, as it was developed with an organizational performance lens that focused on all levels of the organization (Rummler & Brache, 1995). The intervention content focus was categorized as problem solving and decision making; strategic stewardship; employee performance; human relations; job and work redesign; and general management.

Outcome Categories. The most pertinent variable, and the unit of analysis for the study, was the outcome resulting from each intervention. A combination of the Results Assessment System (Swanson & Holton, 1999) and Burke and Day's (1986) outcomes were used to define the outcomes of leadership development studies from both a learning and a performance perspective.

Outcome categories were defined as having either an objective or subjective (Burke & Day, 1986) and either performance- or learning-level outcome (Swanson & Holton, 1999). Performance-level outcomes were further categorized as system results, and learning-level outcomes were delineated into knowledge or expertise (behavior) results. Perception outcomes were not included. The outcome categories were adapted from Swanson and Holton (1999) and Burke and Day (1986) and were specifically defined for this research as follows:

Knowledge—Subjective: Principles, facts, attitudes, and skills learned during or by the end of training as communicated in statements of opinion, belief, or judgment completed by the participant or trainer

Knowledge—Objective: Principles, facts, attitudes and skills learned during or by the end of training by objective means, such as number of errors made or number of solutions reached, or by standardized test

Behavior/Expertise—Subjective: Measures that evaluate changes in on-the-job behavior perceived by participants, or global perceptions by peers or a supervisor

Behavior/Expertise—Objective: Tangible results that evaluate changes in on-the-job behavior or supervisor ratings of specific observable behaviors

System Results/Performance—Subjective: Organization results perceived by respondents, not reported by company records (for example, subordinates' job satisfaction or commitment to the organization), and group effectiveness perceived by subordinates.

System Results/Performance—Objective: Tangible results, such as reduced costs, improved quality or quantity, promotions, and reduced number of errors in making performance ratings

Categorization by Research Designs. When a meta-analyst is faced with multiple research designs, he or she has one of two choices: integrate all research design types to create one overall effect size, or conduct separate meta-analyses of the studies based upon research design types and create separate effect sizes per research design. According to Hunter and Schmidt (1990), the data from studies with different research designs "must be analyzed in different ways using different formulas for sampling error" (p. 339). While the effect sizes could be statistically aggregated, we decided that the effects represented substantively different criterion (for example, gain scores versus post-training mean versus control group), so we elected to conduct separate meta-analyses for each research design. To conduct individual meta-analyses by the type of research design provides an aggregation of similar studies and the ability to test the research design as a moderator variable.

Unfortunately, single group pretest-posttest (SGPP) studies are often overlooked as a valuable data source in meta-analysis because results are believed to have weak controls and threats to internal and external validity. Some researchers believe that data from SGPP designs "upwardly bias the mean treatment effect

estimates derived from meta-analysis" (Lipsey & Wilson, 1993, p. 1194). Nevertheless, SGPP design was included, as it is often used to evaluate training programs (Carlson & Schmidt, 1999) and measure individual growth and learning. Hunter and Schmidt (1990) demonstrated that under most circumstances "the within-subjects design is far superior to the between-subjects design" (p. 339) and "has a much higher statistical power than does the independent groups subjects design" (p. 341). They suggest that when a treatment by subject interaction is present that is always true of training (that is, the individual differences of the participants interacts with the intervention), then SGPP may be the best design. Also, Hunter and Schmidt (1990) "urged experimenters to use the more powerful within-subjects design whenever possible" (p. 340). SGPP is a within-subjects design. For high-level executive leadership development, SGPP may be the only design possible, since a control group cannot be obtained. Therefore, it is important to examine the SGPP studies.

This research is unique as separate meta-analyses were conducted by research design: posttest only with control group (POWC); pretest-posttest with control group (PPWC); correlation (CORR); and single group pretest-posttest (SGPP). Each of the four research designs had the potential of having seven outcome subgroups: knowledge objective and subjective; expertise objective and subjective; and system objective and subjective. Thus twenty-eight potential outcome subgroups existed. In the final research sample, we used outcome subgroups with six or more effect sizes. While a meta-analysis can be performed with a sample as small as two effect sizes, meta-analytic researchers (Sackett, in press) caution against drawing strong conclusions from such a small sample of studies. Power depends on the conditions simulated, the Type I error rate, and the magnitude of validity difference one is interested in detecting. Therefore, no single value can be used to represent power.

Space limitations prohibit a full discussion of this, but the thirty-two potential outcome subgroups were reduced to eleven once subgroups with fewer than six studies were dropped. In two cases, we were able to preserve some data by dropping the pretest-posttest comparison for the pretest-posttest with control (PPWC) knowledge objective and system objective outcome subgroups and integrating the studies with posttest only with control (POWC) knowledge objective and system objective outcome subgroups, respectively. Correlational studies were dropped because too few studies were available to conduct a meta-analysis with significance. The final sample of managerial leadership development studies was reduced to eighty-three. The posttest only with control (POWC) sample contained thirty-six studies, the pretest-posttest with control (PPWC) sample contained twenty-six studies, and the single group pretest-posttest (SGPP) sample contained twenty-five studies.

Effect Size Calculations. The key dependent variable was the effect size, a statistic that encoded the critical quantitative information from each relevant study finding to transform data from each study into a common metric (Lipsey & Wilson, 2001). Carlson and Schmidt's (1999) formulas were used

to determine effect sizes and reduce the amount of sampling error associated with the effect size estimate. Carlson and Schmidt refined Hunter and Schmidt's (1990) effect size formulas, allowing for single group pretest-posttest research designs to be included in meta-analytic research. To control for some of the threats to validity, Carlson and Schmidt's (1999) formulas used pretest standard deviation instead of posttest standard deviation to calculate the effect size change. For consistency purposes, Carlson and Schmidt's formulas were used to determine effect sizes in each of the three meta-analyses in this research.

The effect size for posttest only with control (POWC) studies was the difference between the mean posttest scores of the treatment and control groups divided by the pooled standard deviation of the two groups. The POWC effect was the normalized difference between a trained and untrained group.

The pretest-posttest with control (PPWC) effect size was the normalized difference in the gain scores between the trained group and the untrained comparison group. The PPWC effect size formula subtracted the raw mean difference of the control group pretest-posttest scores from the raw mean difference of the treatment pretest-posttest scores and divided the resultant average gain by the pooled standard deviation of the training and control groups pre-training dependent variable assessments. Of important note was the use of the pre-training standard deviation in calculating both PPWC and single group pretest-posttest (SGPP) effect sizes, as post-training can be altered by individual differences that interact with training methods employed, resulting in participants learning at different rates (Glass, McGaw, & Smith, 1981). These differences may be due to inattention or distractions in the learning environment, differing opportunities for participation, or differential exposure to treatment.

The effect size for single group pretest-posttest (SGPP) studies was determined by comparing the pretest and posttest mean scores for the trained group and dividing the resultant comparison by the standard deviation of the pre-training dependent variable measure (Hunter & Schmidt, 1990).

In studies where the mean, standard deviation, *t*-value, *p*-value and the standard difference (*d*) were also reported, the means and standard deviation served as the primary set of statistics from which to determine an effect size, followed by *t*-value when available. The *p*-value was obtained from the table of probabilities associated with observed values of *z* in the normal distribution (Hunter & Schmidt, 1990).

Unit of Analysis. The unit of analysis was the outcome of each intervention. When more than one measure of a dependent variable was used, the resulting data was weighted to produce one effect size per dependent variable. For example, if expertise objective outcomes were measured by two different measures in a study, these two measures were weighted by group size to produce one effect size for expertise objective outcomes. When multiple independent interventions occurred within the same study, we employed Hunter, Schmidt, and Jackson's (1982) design, where they were considered to be

separate studies that used the same research designs and tests. We also employed Hunter and Schmidt's (1990) recommendation that the multiple outcomes be treated as if they were independent studies, with effect sizes computed for each intervention.

Correction for Statistical Artifacts. This research corrected for two artifacts: sampling error and error of measurement. Artifacts not controlled have the effect of lowering the observed effect size (Aguinis & Pierce, 1998; Hunter & Schmidt, 1990). A weighted effect size was calculated based upon the number of subjects in the intervention so that studies of different sizes or more individuals subjected to the intervention were not treated as though they made the same contribution to the conclusions. The mean and variance of validity coefficients were also corrected for attenuation due to measurement error by aggregating all measurement reliability coefficients and applying a global correction method, adjusting the mean values over all effect sizes, even when each individual effect size could not be determined from the study. Studies were not corrected for range restriction; there was no reason a priori to suspect a restricted range, since the participants in the interventions met the population of interest (supervisors, managers, and leaders in organizations). A search was made of the Forrest plot for extreme effect sizes (that is, two or more standard deviations beyond the mean of its respective group), and no outliers were found.

Moderator Analysis. Meta-analysis research provides for testing differences in across-study variability in effect size for underlying phenomena or moderators. Two groups of potential moderators were identified: those within a particular research design and those spanning all studies. Five potential moderating variables were identified within the first group: the content focus of the intervention, organization type, measurement method, job classification, and publication type. However, the moderator analysis results could not be utilized with any reasonable level of confidence for two reasons: the low power of the studies and the probability of experiment-wise error. First, in many of the moderator-outcome combinations in this meta-analysis, there were subgroups analyzed that had a very small number of effect sizes, due primarily to the division of the studies into different research design groups. In many subgroups, only one or two effect sizes existed, which may mean that the power was too low to detect all moderator effects in those combinations. Small numbers of studies also generate a distribution of effect sizes, which is what is expected in effective meta-analytic research. Thus, any one of the moderator effects presented in this research could possibly be an artifact, due to the small number of effect sizes in subgroups of the moderator variable.

Second, with a 5 percent probability of finding a moderator, in fifty-four moderator-dependent variable combinations present in this study, it is likely that three cell combinations would have moderators. Seven cell combinations in this meta-analysis were identified as possibly having a moderator. Thus, one could expect that approximately one-half of the moderators found would occur through random sampling, making the overall impact of this moderator analysis

even more suspect. Thus, the results from this group of moderators are *not* reported here.

Two potential moderator variables were identified in the second group, spanning all studies, and are reported in this study: research design type and subjectivity-objectivity. Once effect sizes were adjusted for artifacts, Hunter and Schmidt's (1990) preferred method of breaking the data into subgroups was used to determine whether a study characteristic was a moderating variable. This approach uses a Q-statistic ratio to partition the observed effect size variability into two components: the portion attributable to subject-level sampling error and the portion attributable to other between-study differences. The distribution was considered to be homogeneous when the sampling error accounted for 75 percent or more of the observed variability (Hunter & Schmidt, 1990). The variable was considered to be a moderator when the residual variance of the population was 25 percent or more of the observed variance. For computational purposes, when the Q value for between groups was more than 25 percent of the total Q value for the subgroup, the grouping variable was considered to be a moderator.

Software for Analysis. All data from the coding forms were entered into *Comprehensive Meta-Analysis, Version 1.0.23,* a software program generated by Borenstein and Rothstein (1999) of Biostat, Inc. This software uses Hunter and Schmidt's (1990) methodology, which allowed for synthesis of data from multiple studies and provided a means for tracking the source of variation when effect sizes differed significantly.

Results

One hundred and three leadership development studies with a full range of managerial leadership development interventions were initially located. Eighty percent were formal training interventions (eighty-three studies), 13 percent (thirteen studies) were feedback interventions, 5 percent (five studies) were coaching and mentoring, and 2 percent were on-the job interventions (two studies). Studies were published primarily in psychology and business or management sources. The majority of interventions had behavioral outcomes and a human relations or general management training content focus. There appeared to be a trend toward multiple training techniques, a blend of cognitive knowledge and behavioral learning, and multiple evaluation techniques that included evaluations by supervisors, subordinates, and peers, along with self-assessments.

There was an average of 1.8 effect sizes per study. Fifty-two studies (40 percent) had expertise subjective outcomes and contributed sixty-three effect sizes. Forty-five studies (34 percent) with expertise objective outcomes produced fifty-two effect sizes. Eleven studies measured system objectives, and only one study had financial outcomes. Twenty-two studies (17 percent) measured knowledge outcomes and contributed twenty-four effect sizes, and

ninety-seven studies had interventions with expertise outcomes, producing 115 (75 percent) of the effect sizes. Forty-six different measurement instruments were used. Cronbach's alpha coefficients were provided for forty-five studies, with an average reliability over those studies of .893.

Posttest Only with Control (POWC) Meta-Analysis. Thirty-six POWC studies generated fifty-nine effect sizes, with a total of 3,104 subjects as indicated in Table 1.

Eleven studies with knowledge objective outcomes produced thirteen effect sizes from 877 subjects, with an overall effect size of .96. Thirteen studies contributed sixteen effect sizes and 843 subjects in the expertise objective subgroup, producing an overall effect size of .54. Eighteen studies contributed twenty-three effect sizes from 966 subjects in the expertise-subjective outcome subgroup, with an average effect size of .41. Seven studies with system objective outcomes produced an overall effect size of .39 from seven effect sizes and 418 subjects. The effect sizes varied among individual studies in the POWC data set from −1.39 to 2.02.

Pretest-Posttest with Control (PPWC) Meta-Analysis. A total of twenty-six studies generated forty-one effect sizes from a total of 1,573 subjects in the PPWC meta-analysis, as shown in Table 2, which presents effect sizes per outcome subgroup for PPWC studies.

Nineteen studies with 1,104 subjects contributed twenty-two effect sizes with expertise objective outcomes, producing an overall effect size of .35. Eighteen studies with 469 subjects contributed nineteen effect sizes and an overall effect size of .40 for expertise subjective outcomes. Eleven studies had more than one type of dependent variable, and two studies had more than one experimental treatment. The effect sizes varied among individual studies in the PPWC data set from −.45 to 1.67.

Single Group Pretest-Posttest (SGPP) Meta-Analysis. The aggregation of SGPP studies was based on twenty-five studies that generated thirty-five effect sizes. The effect sizes varied among individual studies in the SGPP data set from −.26 to 2.10, as shown in Table 3.

Six studies produced six effect sizes in the knowledge objective subgroup, with an effect size of 1.37 from 642 subjects. Fourteen studies with fourteen effect sizes and 1,004 subjects with expertise objective outcomes produced an effect size of 1.01. Thirteen studies with fifteen effect sizes and 2,638 subjects in the expertise subjective outcome subgroup produced an effect size of .38. Twelve studies had more than one dependent variable, and two studies had more than one experimental treatment.

Discussion and Conclusions

This research differs from other meta-analyses because studies from all research designs were located and reviewed for inclusion. Single group pretest-posttest measurements without control groups (SGPP) were included, because they are

Table 1. Effect Sizes of Interventions per Outcome Subgroup (POWC)

Citation (Year)	*N (Training)*	*N (Control)*	*Effect Size*	*SE*	*95 Percent Confidence Interval* *Lower Limit*	*Upper Limit*	*Org Type*	*Job Class*	*Content*
Knowledge-Objective									
Davis/Mount (1984)	88	122	1.60	.16	1.28	1.91	O	M	E
Davis/Mount (1984)	135	122	1.19	.14	.92	1.45	O	M	E
DeNisi/Peters (1996)	66	22	.65	.25	.15	1.15	Ma	E	H
Haccoun/Hamtiaux (1994)	42	24	.83	.27	.29	1.36	E	M	H
Harrison (1992)	11	12	1.44	.47	.46	2.41	Mi	O	H
Harrison (1992)	11	12	1.56	.48	.57	2.55	Mi	O	H
Maurer/Fay (1988)	212	21	1.10	.23	.63	1.56	Me	O	H
May/Kahnweiler (2000)	19	19	1.01	.34	.31	1.71	Ma	E	H
Russell et al. (1984)	19	11	1.39	.42	.53	2.25	A	M	H
Thoms/Klein (1994)	64	64	.32	.27	−.18	.85	Me	O	H
Tziner et al. (1991)	45	36	.72	.23	.26	1.17	Mi	E	H
Wolf (1996)	144	144	.76	.12	.52	1.00	Me	O	H
Yaworsky (1994)	21	21	.61	.32	−.03	1.25	G	E	G
Subtotal ($n = 11$; $k = 13$)	877	630	.96	.07	.82	1.12			
Expertise-Objective									
Alsamani (1997)	38	31	1.39	.27	.85	1.93	G	M	G
Bendo (1984)	63	63	.39	.18	.03	.74	O	E	H
Clark et al. (1985)	19	8	.84	.44	−.06	1.74	Me	O	H
Davis/Mount (1984)	66	96	−1.39	.18	−1.75	−1.05	O	M	E
Davis/Mount (1984)	104	98	.40	.14	.12	.68	O	M	E
DePiano/McClure (1987)	23	60	.74	.25	.23	1.24	E	T	H
Dvir et al. (2002)	23	17	−.90	.34	−1.56	−.22	Mi	E	S
Earley (1987)	20	20	2.01	.39	1.22	2.79	Ma	E	G
Earley (1987)	20	20	1.03	.34	.34	1.71	Ma	E	G
Earley (1987)	20	20	1.44	.35	.72	2.16	Ma	E	G
Eden (1986)	7	9	.89	.53	−.24	2.02	Mi	O	J
Eden et al. (2000)	138	184	.06	.11	−.16	.28	Mi	O	H
Henry (1983)	156	156	.24	.11	.01	.46	T	E	H
Josefowitz (1984)	65	31	−.08	.22	−.52	.35	T	E	G
Scandura/Graen (1984)	21	57	.51	.26	−.01	1.02	G	E	H
Thoms/Klein (1994)	60	60	.41	.29	−.11	.96	Me	O	H
Subtotal ($n = 13$; $k = 16$)	843	870	.54	.19	.14	.95			
Expertise-Subjective									
Alsamani (1997)	38	31	1.00	.26	.49	1.51	G	M	G
Briddell (1986)	24	24	.02	.29	−.56	.60	E	O	G
Colan/Schneider (1992)	184	50	.26	.16	−.05	.58	U	E	H
Dvir et al. (2002)	27	18	.52	.31	−.11	1.14	Mi	E	S
Earley (1987)	20	20	1.23	.34	.53	1.93	Ma	E	G
Earley (1987)	20	20	1.15	.34	.46	1.84	Ma	E	G
Earley (1987)	20	20	1.53	.36	.80	2.26	Ma	E	G
Eden (1986)	7	9	.89	.53	−.24	2.02	Mi	O	J
Fuller (1985)	24	24	−.48	.29	−1.07	.11	E	M	H

(*Continued*)

Table 1. Effect Sizes of Interventions per Outcome Subgroup (POWC) (*Continued*)

					95 Percent Confidence Interval				
Citation (Year)	*N (Training)*	*N (Control)*	*Effect Size*	*SE*	*Lower Limit*	*Upper Limit*	*Org Type*	*Job Class*	*Content*
Gerstein et al. (1989)	112	112	.23	.13	−.03	.50	A	O	G
Harrison (1992)	11	12	.62	.43	−.27	1.50	Mi	O	H
Harrison (1992)	11	12	.02	.42	−.85	.88	Mi	O	H
Henry (1983)	156	156	.28	.11	.06	.51	T	E	H
Ivancevich (1992)	15	15	.62	.37	−.14	1.39	O	O	H
Ivancevich (1992)	15	15	.74	.38	−.04	1.51	O	O	H
Ivancevich (1992)	15	15	.43	.37	−.33	1.18	O	O	H
Maurer/Fay (1988)	21	21	−.44	.31	−1.07	.19	Me	O	H
Moxnes/Eilerten (1991)	77	133	.14	.14	−.14	.42	U	E	G
Reaves (1993)	25	20	1.01	.32	.37	1.65	E	O	H
Scandura/Graen (1984)	21	57	.38	.26	−.13	.89	G	E	H
Tziner et al. (1991)	45	36	.56	.23	.10	1.01	Mi	E	H
Williams (1992)	49	30	.42	.23	−.32	.60	G	E	G
Young/Dixon (1995)	29	40	−.78	.25	.27	1.28	O	T	G
Subtotal (n = 18; k = 23)	966	890	.41	.22	.25	.58			
System-Objective									
Bankston (1993)	13	13	.79	.31	.42	.93	E	T	H
Colan/Schneider (1992)	184	50	.30	.16	−.01	.62	U	E	H
Graen et al. (1982)	36	95	.60	.20	.20	.99	G	O	J
Hill (1992)	25	27	.10	.27	−.46	.66	E	T	H
Posner (1982)	34	34	.02	.24	−.46	.50	G	E	H
Scandura/Graen (1984)	21	57	.67	.26	.15	1.19	G	E	H
Urban et al. (1985)	105	105	.43	.14	.16	.71	O	E	G
Subtotal (n = 7; k = 7)	418	381	.39	.10	.19	.59			
Combined (n = 36; k = 59)	3,104	2,771							

Note: Organization Type (Org Type): G = Government; Ma = Manufacturing; T = Technology; E = Education; Mi = Military; U = Utilities; Me = Medical; A = Automotive; and O = Other Unknown. Job Classification (Job Class): T = Top Management; M = Mid-Manager; E = Entry Level; and O= Mixed Groups. Training Content (Content): G = General Management; H = Human Relations; S = Strategic Stewardship; E = Employee Performance; and J = Job and Work Redesign. n = number of studies; k = effect size; N = number of subjects.

often used to evaluate training programs (Carlson & Schmidt, 1999) and to measure individual growth and learning. Not to include these studies would leave out a significant group of managerial leadership development interventions. Overall, the effectiveness of managerial leadership development programs varied widely: some programs were tremendously effective, and others failed miserably. For example, effect sizes in the individual studies in

Table 2. Effect Sizes of Interventions per Outcome Subgroup (PPWC)

Citation (Year)	*N (Training)*	*N (Control)*	*Effect Size*	*SE*	*95 Percent Confidence Interval* *Lower Limit*	*Upper Limit*	*Org Type*	*Job Class*	*Content*
Knowledge-Objective									
Bankston (1993)	13	15	.57	.38	−.23	1.36	E	T	H
Birkenbach et al. (1984)	25	25	.89	.30	.29	1.48	Ma	E	H
Deci et al. (1989)	235	177	.49	.10	.29	.69	O	O	H
Devlin-Scherer et al. (1997)	162	162	.04	.11	−.18	.25	E	O	H
Eden et al. (2000)	17	17	1.29	.38	.52	2.05	E	T	H
Eden et al. (2000)	12	10	.27	.43	−.62	1.17	F	T	H
Frost (1996)	33	17	.37	.30	−.24	.97	G	O	H
Frost (1996)	28	17	−.45	.31	−1.07	.18	G	O	H
Graen et al. (1982)	37	95	.74	.20	.35	1.14	G	O	J
Graen et al. (1982)	37	95	.51	.20	.11	.89	G	O	J
Mattox (1985)	18	18	−.05	.33	−.73	.63	O	O	G
May/Kahnweiler (2000)	19	19	1.01	.35	.31	1.71	Ma	E	H
Nelson (1990)	30	14	.04	.32	−.61	.69	E	T	G
Niska (1991)	13	13	1.12	.42	.25	2.00	E	T	G
Rosti/Shipper (1998)	27	26	.41	.28	−.15	.97	O	M	G
Russell et al. (1984)	9	16	.23	——	−.64	1.09	A	M	H
Savan (1983)	25	25	.07	.28	−.50	.64	O	E	H
Smith et al. (1992)	14	14	1.22	.41	.38	2.07	E	O	H
Sniderman (1992)	59	146	.20	.20	−.19	.59	O	O	H
Steele (1984)	220	219	.03	.10	−.16	.22	T	M	H
Tharenou/Lyndon (1990)	50	50	.78	.21	.37	1.19	G	E	G
Yaworsky (1994)	21	21	.09	.31	−.54	.71	G	E	G
Subtotal ($n = 19$; $k = 22$)	1,104	1,211	.35	.08	.20	.50			
Expertise-Subjective									
Bankston (1993)	13	15	.33	.33	−.45	1.11	E	T	H
Barling et al. (1996)	9	11	.42	.45	−.53	1.38	F	T	S
Birkenbach et al. (1984)	25	25	.78	.29	.19	1.37	Ma	E	H
Cato (1990)	40	40	.29	.22	−.16	.74	G	O	G
Clark (1990)	31	63	.11	.22	−.33	.54	E	E	H
Deci et al. (1989)	8	13	.20	.45	−.74	1.15	O	O	H
Eden et al. (2000)	17	17	.60	.35	−.11	1.31	E	T	H
Eden et al. (2000)	21	23	.57	.31	−.05	1.20	F	T	H
Edwards (1992)	29	39	1.22	.27	.69	1.76	E	O	P
Hill (1992)	25	27	−.13	.28	−.69	.42	E	T	H
Lawrence/Wiswell (1993)	33	32	.29	.25	−.20	.79	G	O	G
Mattox (1985)	18	18	−.15	.33	−.83	.53	O	O	G
Nelson (1990)	30	14	.17	.32	−.48	.82	E	T	G
Niska (1991)	13	13	1.67	.46	.73	2.62	E	T	G
Russell et al. (1984)	11	17	−.22	.38	−1.01	.58	A	M	H
Savan (1983)	25	25	−.01	.28	−.57	.56	O	E	H
Steele (1984)	50	50	.09	.20	−.30	.49	T	M	H

(*Continued*)

Table 2. Effect Sizes of Interventions per Outcome Subgroup (PPWC) (*Continued*)

					95 Percent Confidence Interval				
Citation (Year)	*N (Training)*	*N (Control)*	*Effect Size*	*SE*	*Lower Limit*	*Upper Limit*	*Org Type*	*Job Class*	*Content*
Tharenou/Lyndon (1990)	50	50	1.06	.21	.64	1.49	G	E	G
Yaworsky (1994)	21	21	.09	.31	−.54	.71	G	E	G
Subtotal (n = 18; k = 19)	469	417	.40	.10	.20	.61			
Combined (n = 26; k = 41)	1,573	1,607							

Note. Organization Type (Org Type): G = Government; Ma = Manufacturing; T = Technology; E = Education; Mi = Military; F = Financial; A = Automotive; and O = Other Unknown. Job Classification (Job Class): T = Top Management; M = Mid-Manager; E = Entry Level; and O = Mixed Groups. Training Content (Content): G = General Management; H = Human Relations; S = Strategic Stewardship; J = Job and Work Redesign; and P = Problem Solving. n = number of studies; k = effect size; N = number of subjects.

this meta-analysis ranged from −1.39, representative of a highly unsuccessful program, to a successful program with an effect size of 2.10.

Studies included in this meta-analysis span the twenty-year period after the legendary Burke and Day (1986) study that is commonly regarded as the principal empirical support for the effectiveness of managerial training and leadership development programs. Table 4 presents a comparison of this research with Burke and Day's meta-analysis (1986) on the effectiveness of managerial leadership development programs. The researchers found methodological differences and suggest that comparisons to Burke and Day's (1986) meta-analysis results should be made with caution. The unit of analysis in this research was a single outcome measure, whereas Burke and Day (1986) calculated an effect size for each dependent variable within a single study. Burke and Day aggregated studies involving 3,967 treated and 3,186 control group participants, but reported 46,574 total subjects across 472 effect sizes. This research reported 9,590 total subjects across 136 effect sizes in all three research designs. Burke and Day's (1986) methodology raises two issues: (1) independence of outcomes measured (effect sizes), and (2) over-weighting of studies with multiple effect sizes. Burke and Day's procedure potentially introduces substantial error, as the inflated sample size, the distortion of standard error estimates arising from the inclusion of non-independent effect sizes, and "the overrepresentation of those studies that contribute more effect sizes can render the statistical results highly suspect" (Lipsey & Wilson, 2001, p. 105). To be fair, meta-analysis methods were still developing at the time they conducted their study, and their method was accepted at that time. However, newer meta-analysis methods are now available that raise questions about the older methods.

Also, Burke and Day (1986) combined all behavioral outcomes into a *subjective behavior* subgroup, and all results outcomes into an *objective results*

Table 3. Effect Sizes of Interventions per Outcome Subgroup (SGPP)

				95 Percent Confidence Interval				
Citation (Year)	*N (Training)*	*Effect Size*	*SE*	*Lower Limit*	*Upper Limit*	*Org Type*	*Job Class*	*Content*
Knowledge-Objective								
Couture (1987)	13	1.06	.42	.19	1.92	Me	E	H
Larsen (1983)	9	1.22	.51	.13	2.31	Me	M	G
Martineau (1995)	67	.66	.18	.31	1.01	O	E	H
Tesoro (1991)	99	1.59	.16	1.27	1.91	O	O	H
Tracey et al. (1995)	104	1.66	.16	1.34	1.97	O	E	G
Yang (1988)	350	3.54	.09	1.37	1.71	G	O	G
Subtotal ($n = 6$; $k = 6$)	642	1.36	.08	1.18	1.56			
Expertise-Objective								
Bruwelheide/Duncan (1996)	12	1.43	.46	.48	2.38	O	O	H
Donohoe et al. (1997)	89	1.66	.17	1.32	2.00	G	O	H
Dorfman et al. (1986)	121	.09	.13	−.16	.34	E	E	E
Jalbert et al. (2000)	39	−.28	.23	−.73	.18	T	T	G
Katzenmeyer (1988)	50	.41	.20	.01	.81	E	T	H
Larsen (1983)	12	.85	.40	.03	1.67	Me	M	G
McCauley/Hughes-James (1994)	38	.25	.23	−.21	.71	E	T	G
Paquet et al. (1987)	22	.60	.31	−.02	1.22	O	O	G
Shipper/Neck (1990)	10	.78	.46	−.19	1.76	Me	O	H
Tesoro (1991)	11	.44	.43	−.46	1.34	O	O	H
Tracey et al. (1995)	104	1.25	.15	.95	1.55	O	E	G
Warr/Bunce (1995)	106	.16	.14	−.11	.43	O	E	G
Woods (1987)	40	.06	.22	−.38	.51	E	O	H
Yang (1988)	350	1.39	.08	1.23	1.56	G	O	G
Subtotal ($n = 14$; $k = 14$)	1,004	1.01	.06	.87	1.15			
Expertise-Subjective								
Faerman/Ban (1993)	1,363	.28	.04	.21	.36	G	E	H
Innami (1994)	112	1.71	.16	1.40	2.02	Me	E	P
Innami (1994)	112	.29	.13	.02	.55	Me	E	P
Katzenmeyer (1988)	50	.69	.21	.28	1.10	E	T	H
Lafferty (1998)	233	.34	.09	.16	.53	Mi	M	G
Lafferty (1998)	282	.29	.08	.12	.45	Mi	M	G
Larkin (1996)	23	.87	.31	.25	1.49	Me	O	H
Martineau (1995)	52	.39	.20	−.01	.78	O	E	H
McCauley/Hughes-James (1994)	38	.61	.23	.14	1.08	E	T	G
Robertson (1992)	160	.04	.11	−.18	.26	O	O	G
Sogunro (1997)	29	2.10	.33	1.44	2.75	O	O	G
Tenorio (1996)	19	.84	.34	.15	1.52	T	E	G
Thoms/Greenberger (1998)	105	.52	.17	.17	.86	O	O	S
Werle (1985)	20	.56	.32	−.09	1.22	O	M	G
Woods (1987)	40	.34	.23	−.11	.79	E	O	H
Subtotal ($n = 13$; $k = 15$)	2,638	.38	.04	.30	.46			
Combined ($n = 25$; $k = 35$)	4,284							

Note: Organization Type (Org Type): G = Government; T = Technology; E = Education; Mi = Military; Me = Medical; and O = Other Unknown. Job Classification (Job Class): T = Top Management; M = Mid-Manager; E = Entry Level; and O = Mixed Groups. Training Content (Content): H = Human Relations; E = Employee Performance; P = Problem Solving; G = General Management; and S = Strategic Stewardship. n = number of studies; k = effect size; N = number of subjects.

Table 4. Comparison of Meta-Analyses on Managerial Leadership Development Programs by Outcome Subgroup

Outcome Subgroup	*Burke and Day (1986)*	*Collins POWC (2002)*	*Collins PPWC (2002)*	*Collins SGPP (2002)*
Knowledge				
Objective	.38	.96	—	1.37
Subjective	.34	—	—	—
Expertise				
Objective	—	.54	.35	1.01
Subjective	.49	.41	.40	.38
System				
Objective	.67	.39	—	—
Subjective	—	—	—	—

Note: Burke and Day (1986): 70 studies, 472 effect sizes (3,967 subjects); Collins POWC (2002): 36 studies, 59 effect sizes (3,104 subjects); Collins PPWC (2002): 26 studies, 42 effect sizes (1,573 subjects); Collins SGPP (2002): 25 studies, 35 effect sizes (4,284 subjects).

subgroup. This research mixed neither objective nor subjective nor financial and system outcomes. Thus, a problem exists of comparing "apples and oranges" when comparing effect sizes. It is unreasonable to believe that the standard error for financial returns would be the same as the standard error for system outcomes. To combine outcome subgroups is to assume that the outcomes are equivalent with equivalent distributions.

In addition, Burke and Day aggregated studies only from business and industry, whereas the current sample included studies from education, government, medical, and military. Of key importance is that two studies in Burke and Day's (1986) meta-analysis and eleven studies in this overall meta-analysis research had organizational-level outcomes. Managerial leadership development is a young field for which little is reported in the literature regarding what is or what is not effective, particularly relative to outcomes of the organization as a system. The difference in effect sizes is very curious. One would expect some differences, because Burke and Day (1986) limited their research to business and industry, which caused the overall focus to be somewhat different.

Knowledge (Learning) Outcomes. Learning outcomes remain a primary focus of leadership development programs. Posttest only with control (POWC) knowledge objective outcomes measured primarily by knowledge tests are highly effective, almost one standard deviation higher than the untrained control group. It would stand to reason that managerial ranks would know why they needed the information provided in training, and could understand why it would be of benefit to them in their own positions. This result compares with .38 effect size found by Burke and Day (1986), indicating only moderate effectiveness. Single group pretest-posttest (SGPP) knowledge objective results had an average effect of 1.37, but a limited number of studies existed on which to base any conclusions. One could assume that the effect would be higher in SGPP studies, where individual differences are recognized and treatment by subject interaction is incorporated and measured. However,

this area could use further research, and especially with SGPP studies, in regard to treatment by subject interaction.

It should be noted that the effect size is higher for knowledge outcomes and gradually dropped for expertise and system across different designs. One would anticipate that top-level leaders would score well on knowledge as measured objectively primarily through testing. Behavior, on the other hand, is much more difficult to measure, especially rated subjectively, and it should be noted that the majority of the studies had expertise outcomes. Because the single group pretest-posttest design (SGPP) captures treatment by subject interaction, the researchers believe the SGPP effect size (1.01) to be the best representation of a behavior effect size. It must also be observed that the effect size did not drop as much between knowledge and behavior for SGPP as in other designs. The effect size of system-level outcomes, on the other hand, was determined with a sample of seven studies, considered to be a small meta-analysis. Inadequate evaluation methods to measure organizational performance and the small sample size lead the authors to speculate on the validity of the effect size for system outcomes until more empirical studies are incorporated.

Expertise Objective Outcomes. The expertise objective outcomes were moderately effective across the posttest only with control (POWC) and pretest-posttest with control (PPWC) meta-analyses, and highly effective in single group pretest-posttest (SGPP) studies. The most obvious measure of behavioral change (expertise) is gain scores, the difference between individual's performance before training (pretest score) and performance after completion of the training (posttest score), as represented in SGPP studies. Behavior change when measured objectively from pretest to posttest was approximately 1.0 standard deviation greater after training. Adult learning principles alert us that individuals react differently to training based on their differences, a concept known as *treatment by subject interaction.* Managers who participate in training programs differ greatly, and for programs to be effective they must accommodate individual managers' abilities, learning styles, and preferences. For example, some people learn best from lectures, others from structured exercises or direct experiences. Pretest-posttest designs are the only ones that incorporate the effect of treatment by subject interaction. The SGPP effect size for expertise objective outcomes (1.01) may be the most reflective of the true effect size. The POWC studies, which are more directly comparable to Burke and Day (1986), showed an effect size of .54, close to their finding of .49.

Expertise Subjective Outcomes. Expertise subjective outcomes were found to have a moderate effectiveness in posttest only with control (POWC), pretest-posttest with control (PPWC), and single group pretest-posttest (SGPP) studies. The outcome for the trained group in POWC expertise subjective studies was .41 standard deviation higher than the control group. In PPWC expertise subjective studies, the behavior change of the trained group was .40 standard deviation higher than the untrained group. SGPP expertise subjective training participants performed .38 standard deviation greater after training than prior to training.

For the most part, expertise subjective and expertise objective outcomes indicated moderate effectiveness, surprisingly similar to Burke and Day's (1986) results, except for a 1.01 effect size for single group pretest-posttest (SGPP) expertise objective outcomes. The SGPP expertise objective effect size was higher than posttest only with control (POWC) expertise objective. While additional research should be performed to reach a viable conclusion, it is believed by the researchers that SGPP may be a stronger design that detects and measures treatment by subject interaction as discussed earlier (Hunter & Schmidt, 1990). Interestingly, the SGPP expertise objective effect size was also significantly higher than SGPP expertise subjective effect sizes (1.01 objective versus .38 subjective). Objective measurements were primarily multiple ratings of the participant's behavior after training. This finding was surprising because subjective ratings are usually higher than objective ratings. It is possible that self-raters do not see change in themselves as quickly as is detected by supervisors or subordinates. Also, some people are overly critical of themselves and may not rate themselves as highly as others would do. Expertise subjective outcomes were primarily measured by self-assessments.

A larger effect size for single group pretest-posttest (SGPP) studies is not unusual. Leddick's (1987) meta-analysis of training effectiveness using both SGPP and posttest only with control (POWC) studies found that "effects were smaller when true controls or non equivalent control groups were used" (p. 98). Leddick obtained an effect size of .98 for a mixed group of expertise subjective and expertise objective outcomes, very close to the effect size obtained in this SGPP expertise objective outcomes (1.01). Chen's (1994) meta-analysis on the effectiveness of cross-cultural training for managers also found effect sizes from SGPP studies overall higher than with control groups (1.74 versus 1.58).

Combining the two outcome categories into one, as was done by Burke and Day (1986), is unwise, as subjective and objective results are distinct and reflect different measurement strategies. This is apparent by the amount of literature on self versus other measurement, which indicates that self-measurements are usually higher. Therefore, it is strongly believed that findings in this meta-analysis provide a greater understanding about managerial leadership development in terms of behavioral outcomes of training than any other meta-analysis.

System Outcomes. Seven posttest only with control (POWC) system objective studies were located, with an average effect size of .39, ranging from .02 to .79 from only 418 subjects. Thus, performance-level outcomes measured objectively for the trained group was .39 standard deviation higher than the control group. This indicates that interventions with system objective outcomes were moderately effective. In contrast, the average system objective effect size in Burke and Day's study was .67 (2,298 participants).

This meta-analysis shows that there is little research describing strategic stewardship training programs or system-level outcomes that involve transformational leadership primarily at the top management level. This is

unfortunate, as companies must develop the strength to widen perspectives and compete globally in today's dynamic business world. It is perhaps too soon for significant research to occur and be reported regarding strategic stewardship and the need to train leaders in strategy development.

Limitations and Implications for Future Research

One of the glaring problems this research uncovered is that what has been considered the classic study on leadership development effectiveness (Burke & Day, 1986) needs to be updated to utilize modern meta-analytic methods. It is impossible to determine whether studies since 1982 show an improvement, no change, or a decline in effectiveness of leadership development. The data would seem to suggest an improvement in knowledge outcomes, but Burke and Day's methods could well have overweighted some studies with poor outcomes—or vice versa. A complete restatement of their data is needed in order to determine whether there have been changes in intervention effectiveness since 1982.

There has been a resurgence of interest in evaluation of managerial leadership development programs (Alliger, Tannenbaum, Bennett, Traver, & Shotland, 1997; Dionne, 1996; Holton, 1996; Moller & Mallin, 1996), with researchers exploring cause-effect relationships between interventions and the participants' learning, expertise or behavior, and system-level results. However, this research points to some other significant holes in the research that could be promising future research opportunities. First, the small number of reported systematic evaluations of leadership development programs with organizational performance as an outcome (Collins, 2001; Sogunro, 1997) limited our ability to determine a critical effect size for system or financial outcomes. Specifically, less than 10 percent of studies located through this research were focused at the organizational level. While the prevailing principles of most management development literature are rooted in organizational strategy and organizational structure, the relationship between corporate performance and individual leadership still lacks significant empirical support. Because organizations are facing a more competitive global economy with increased performance demands, it is important that evaluation and performance-based management development theory be combined to create the appropriate system for measurement of organizational-level performance. More reported studies with organizational-level outcomes are needed before researchers can determine the overall effectiveness of leadership development programs at the system level. Also, the current research shows there is an emerging trend of transformational leadership, but little training and reporting of results exists in this important area of managerial leadership development.

Second, few empirical studies were available for outcomes of on-the-job assignments, coaching, mentoring, or feedback interventions, which made it impossible to determine the effectiveness of those interventions through this

meta-analysis. These are commonly used managerial leadership development experiences, and are believed to be the interventions at the leading edge of managerial leadership development programs for the future. The literature search also showed encouraging initial evidence of the effectiveness of team training (Eden, 1986; Graen, Novak, & Sommerkamp, 1982). More research and reporting of results is needed to provide definitive use to practitioners regarding the effectiveness of team training, as well as on-the-job experiences and developmental relationships.

Third, small sample sizes limited the evaluation of possible moderators of managerial leadership development interventions. An advantage to meta-analysis is that it allows testing of across-study variables to ensure that no extraneous underlying phenomena or moderators affect the interventions. Additional knowledge regarding moderators is important in our total understanding of the field of leadership development.

Fourth, attention to research design in performing additional meta-analyses could provide a further understanding of leadership development literature. More research and results should be reported from single group pretest-posttest (SGPP) studies, incorporating treatment by subject interaction or the individual learner differences, in response to training. This is important, as SGPP design is often the only type to measure training effectiveness, but is a type of research design often left out of meta-analyses because there is no control group. SGPP measurements should be explored further. Also, the current empirical research also does not maximize the use of the additional data in the studies with a pretest-posttest with control (PPWC) design. PPWC studies should be split into a data set of posttest only with control (POWC) and single group pretest-posttest (SGPP) studies, and results compared.

Fifth, it is important to note that previous meta-analyses do not aggregate studies with expertise objective outcomes. The only known findings of behavior measured objectively are based upon this research. It is suggested that future research separate subjective from objective behavioral outcomes and system outcomes from financial outcomes. This conceptualization should be incorporated to conduct another meta-analysis using the same sample of studies as Burke and Day (1986).

Implications for Practice

Organizations should feel comfortable that their managerial leadership development programs will produce substantial results, especially if they offer the right development programs for the right people at the right time. For example, it is important to know whether a six-week training session is enough or the right approach to develop new competencies that change managerial behaviors, or is it individual feedback from a supervisor on a weekly basis regarding job performance that is most effective?

A wide variety of program outcomes are reported in the literature—some that are effective, but others that are failing. In some respects the lessons for practice can be found in the wide variance reported in these studies. The range of effect sizes clearly shows that it is possible to have very large positive outcomes, or no outcomes at all.

Insufficient data was provided in many studies about the programs, needs assessments, and/or methodology to empirically assess reasons for higher or lower effect sizes. Therefore, as researchers all we can do is speculate on the findings in the current research. Ideally we would like to have coded additional variables from each study in this meta-analysis so that we could establish whether there are other design characteristics that produce better outcomes. For example, we would like to have coded whether or not any types of needs assessments were done. Best practices tell us that an up-front analysis is the key to making sure that the interventions target the right skills to positively impact organizational performance. Actually, the wide variation in effectiveness of the interventions in the studies in this meta-analysis could possibly have been due to poor needs analyses, but because studies often don't report this information, we were unable to determine if this was truly the case. When needs analyses are not done, leadership development programs may incorporate leadership dimensions in the program design that are not appropriate for the organization. Some training professionals make efforts to create favorable conditions for transfer of training and conduct needs assessments so that training objectives address the organization's strategic goals and that resources are directed where they can have the greatest impact on the program and on the participants (Conger & Benjamin, 1999). It helps to develop training objectives that are tailored directly to address the obstacles and dilemmas impacting the implementation of the organization's strategic goals.

This research shows that leadership development has begun to focus on managerial teams and organizational transformation. However, because of the limited number of studies, the relationship between leadership development and organizational performance still remains unclear (Fiedler, 1996; Lynham, 2000). Using the high-performance leadership competency framework (Holton & Naquin, 2000) to conduct future meta-analyses and to develop training programs would integrate multiple leadership perspectives and help researchers be able to more clearly connect organizational performance and leadership development.

The small number of studies in certain outcome categories suggests that organizations must also spend time evaluating the effectiveness of those interventions and reporting the findings. What is not reported in studies in this meta-analysis, or perhaps often overlooked regarding training, is the cost to the organization of trainees in the classroom—the return on investment made by the training program. Large sums of money are invested in managerial leadership development programs annually (Gibler, Carter, & Goldsmith, 2000).

The cost for higher-paid managers to be away from work could be substantial. While it is known that training programs can be effective, organizations should determine the actual return on investment from training initiatives in their organizations.

References

Aguinis, H., & Pierce, C. A. (1998). Testing moderator variable hypotheses meta-analytically. *Journal of Management, 24* (5), 577–592.

Alliger, G. M., & Janak, E. A. (1989). Kirkpatrick's levels of training criteria: Thirty years later. *Personnel Psychology, 42,* 331–342.

Alliger, G. M., Tannenbaum, S. I., Bennett, W., Traver, H., & Shotland, A. (1997). A meta-analysis of the relations among training criteria. *Personnel Psychology, 50* (2), 341–359.

Avolio, B. J. (1999). *Full leadership development: Building the vital forces in organizations.* Thousand Oaks, CA: Sage.

Bass, B. M. (1985). *Leadership and performance beyond expectations.* New York: McGraw-Hill.

Bass, B. M. (1990). *Bass & Stodgill's handbook of leadership: Theory, research, and managerial application.* New York: Free Press.

Bass, B. M. (1998). *Transformational leadership: Industry, military, and educational impact.* Mahwah, NJ: Erlbaum.

Bassi, I. J., Benson, G. W., & Cheney, S. (1996). The top ten trends. *Training and Development, 50,* 18–42.

Bayley, E. W. (1988). *A meta-analysis of evaluations of the effect of continuing education on clinical practice in the health professions.* Unpublished doctoral dissertation, University of Pennsylvania.

Borenstein, M., & Rothstein, H. (1999). *Comprehensive meta-analysis: A computer program for research synthesis.* Englewood, NJ: Biostat.

Brungardt, C. (1996). The making of leaders: A review of the research in leadership development and education. *Journal of Leadership Studies, 3* (3), 81–95.

Burke, M. J., & Day, R. R. (1986). A cumulative study of the effectiveness of managerial training. *Journal of Applied Psychology, 71,* 232–245.

Caligiuri, P. M., & Stroh, L. K. (1995). Multinational corporation management strategies and international human resources practices: Bringing IHRM to the bottom line. *International Journal of Human Resource Management, 6* (3), 494–507.

Campbell, J. A., Dunnette, M. D., Lawler, E. E., & Weick, K. E. (1970). *Managerial behavior, performance, and effectiveness.* New York: McGraw-Hill.

Carlson, K. D., & Schmidt, F. L. (1999). Impact of experimental design on effect size: Findings from the research literature on training. *Journal of Applied Psychology, 84* (6), 851–862.

Chen, L. (1994). *Cross-cultural training effectiveness: A meta-analysis.* Unpublished master's thesis, Central Michigan University.

Clark, K. C., Clark, M. B., & Campbell, D. P. (Eds.). (1992). *Impact of leadership.* Greensboro, NC: Center for Creative Leadership.

Clement, R. W. (1982). Testing the hierarchy theory of training evaluation: An expanded role for trainee reactions. *Public Personnel Management Journal, 11,* 176–184.

Collins, D. B. (2001). Organizational performance: The future focus of leadership development programs. *Journal of Leadership Studies, 7* (4), 43–54.

Collins, D. B., Lowe, J. S., & Arnett, C. R. (2000). In E. F. Holton & S. S. Naquin (Eds.), *Developing high-performance leadership competency, Vol. 6* (pp. 18–46). Baton Rouge, LA: Academy of Human Resource Development.

Conger, J. A., & Benjamin, B. (1999). *Building leaders.* San Francisco: Jossey-Bass.

Dionne, P. (1996). The evaluation of training activities: A complex issue involving different stakes. *Human Resource Development Quarterly, 7* (3), 279–286.

Eden, D. (1986). Team development: Quasi-experimental confirmation among combat companies. *Group and Organization Studies, 11* (3), 133–146.

Fiedler, F. E. (1996). Research on leadership selection and training: One view of the future. *Administrative Science Quarterly, 41,* 241–250.

Fleishman, E. A., Mumford, M. D., Zaccaro, S. J., Levin, K. Y., Korotkin, A. L., & Hein, M. B. (1991). Taxonomic efforts in the description of leader behavior: A synthesis and functional interpretation. *Leadership Quarterly, 2,* 245–287.

Friedman, T. L. (2000). *The lexus and the olive tree.* New York: Anchor Books.

Gibler, D., Carter, L., & Goldsmith, M. (2000). *Best practices in leadership development handbook.* San Francisco: Jossey-Bass.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research.* Thousand Oaks, CA: Sage.

Graen, G., Novak, M., & Sommerkamp, P. (1982). The effects of leader-member exchange and job design on productivity and satisfaction: Testing a dual attachment model. *Organizational Behavior and Human Performance, 30,* 109–131.

Hackman, J. R., & Walton, R. E. (1986). Leading groups in organizations. In P. S. Goodman & Associates (Eds.), *Designing effective work groups* (pp. 72–119). San Francisco: Jossey-Bass.

Herling, R. W. (2000). Operational definitions of expertise and competence. In R. W. Herling & J. Provo (Eds.), *Strategic perspectives of knowledge, competence, and expertise* (pp. 8–21). San Francisco: Berrett-Koehler.

Holton, E. F. III (1996). The flawed four-level evaluation model. *Human Resource Development Quarterly, 7* (1), 5–21.

Holton, E. F. (1999). Performance domains and their boundaries. In R. J. Torraco (Ed.), *Advances in developing human resources, Vol. 1* (pp. 26–46). Baton Rouge, LA: Academy of Human Resource Development.

Holton, E. F., & Naquin, S. S. (2000). *Developing high performance leadership competency.* Baton Rouge, LA: Academy of Human Resource Development.

Hooijberg, R., Hunt. J. G., & Dodge, G. E. (1997). Leadership complexity and development of the leaderplex model. *Journal of Management, 23,* 375–408.

House, R. J., & Aditya, R. N. (1997). The social scientific study of leadership: Quo vadis? *Journal of Management, 23,* 409–473.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis.* Thousand Oaks, CA: Sage.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulating research findings across studies.* Thousand Oaks, CA: Sage.

Huselid, M. A. (1995). The impact of human resource management practices on turnover, productivity, and corporate financial performance. *Academy of Management Journal, 38* (3), 635–672.

Jacobs, R. L., & Jones, M. J. (1995). *Structures on-the-job training: Unleashing employee expertise in the workplace.* San Francisco: Berrett-Koehler.

Kirkpatrick, D. L. (1998). *Evaluating training programs: The four levels.* San Francisco: Berrett-Koehler.

Klenke, K. (1993). Leadership education at the great divide: Crossing into the 21st century. *Journal of Leadership Studies, 1* (1), 112–127.

Kotter, J. (1990). *A forceful change: How leadership differs from management.* New York: Free Press.

Krohn, R. A. (2000). Training as a strategic investment. In R. W. Herling & J. Provo (Eds.), *Strategic perspectives on knowledge, competence, and expertise.* San Francisco: Berrett-Koehler.

Lai, L. C. (1996). *A meta-analysis of research on the effectiveness of leadership training programs.* Unpublished doctoral dissertation, University of Wisconsin-Madison.

Lam, L. L., & White, L. P. (1998). Human resource orientation and corporate performance. *Human Resource Development Quarterly, 9* (4), 351–364.

Larson, C. E., & LaFasto, F.M.J. (1989). *Teamwork: What must go right/What can go wrong.* Thousand Oaks, CA: Sage.

Leddick, A. S. (1987). *Effects of training on measures of productivity: A meta-analysis of the findings of forty-eight experiments.* Unpublished doctoral dissertation, Western Michigan University.

Lepsinger, R., & Lucia, A. D. (1997). *The art and science of 360-degree feedback.* San Francisco: Pfeiffer.

Levi, A. S., & Mainstone, L. E. (1992). Breadth, focus, and content in leader priority-setting: Effects on decision quality and perceived leader performance. In K. E. Clark, M. B. Clark, & D. P. Campbell (Eds.), *Impact of leadership* (pp. 429–442). Greensboro, NC: Center for Creative Leadership.

Lipsey, M. S., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment. *American Psychologist, 48,* 1181–1209.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.

Lynham, S. A. (2000). Leadership development: A review of the theory and literature. In P. Kuchinke (Ed.), *Proceedings of the 2000 Academy of Human Resource Development Annual Meeting.* Baton Rouge, LA: Academy of Human Resource Development.

Marquardt, M. J., & Engel, D. W. (1993). *Global human resource development.* Upper Saddle River, NJ: Prentice Hall.

McCauley, C. D., & Brutus, S. (1998). *Management development through job experiences.* Greensboro, NC: Center for Creative Leadership.

McCauley, C. D., Moxley, R. S., & Van Velsor, E. (1998). *The center for creative leadership handbook of leadership development.* San Francisco: Jossey-Bass.

Moller, L., & Mallin, P. (1996). Evaluation practices of instructional designers and organizational supports and barriers. *Performance Improvement Quarterly, 9* (4), 82–92.

Moxnes, P., & Eilertsen, D. (1991). The influence of management training upon organizational climate: An exploratory study. *Journal of Organizational Behavior, 12,* 399–411.

Newstrom, J. W. (1995). Review of "Evaluating training programs: The four levels of D. L. Kirkpatrick." *Human Resource Development Quarterly, 6,* 417–319.

Rosenthal, R. (1984). *Meta-analysis procedures for social research.* Thousand Oaks, CA: Sage.

Rummler, G. A., & Brache, A. P. (1995). *Improving performance: How to manage the white space on the organizational chart.* San Francisco: Jossey-Bass.

Sackett, P. R. (in press). The status of validity generalization research: Key issues in drawing inferences from cumulative research findings. In K. R. Murphy (Ed.), *Validity generalization: A critical review.* Mahwah, NJ: Erlbaum.

Sogunro, O. A. (1997). Impact of training on leadership development: Lessons from a leadership training program. *Evaluation Review, 21* (6), 713–737.

Swanson, R. A. (1994). *Analysis for improving performance: Tools for diagnosing organizations and documenting workplace expertise.* San Francisco: Berrett-Koehler.

Swanson, R. A. (1998). Demonstrating the financial benefit of human resource development: Status and update on the theory and practice. *Human Resource Development Quarterly, 9* (3), 285–295.

Swanson, R. A., & Holton, E. F. (1999). *Results: How to assess performance, learning, and perceptions in organizations.* San Francisco: Berrett-Koehler.

Ulrich, D. (1997). Measuring human resources: An overview of practice and a prescription for results. *Human Resource Management, 36,* 303–320.

Vaill, P. B. (1990). *Managing as a performing art: New ideas for a world of chaotic change.* San Francisco: Jossey-Bass.

Yukl, G. (1989). Managerial leadership: A review of theory and research. *Journal of Management, 15,* 251–289.

Yukl, G., & van Fleet, D. D. (1992). Theory and research on leadership in organizations. In M. D. Dunnett & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology, Vol. 3* (pp. 147–197). Palo Alto, CA: Consulting Psychologists Press.

Zhang, J. (1999). *Effects of management training on trainees' learning, job performance, and organization results: A meta-analysis of evaluation studies from 1983–1997.* Unpublished doctoral dissertation, Oklahoma State University.

Meta-Analysis Sample

Alsamani-Abdulrahman, S. (1998). *The management development program in Saudi Arabia and its effect on managerial job performance.* Unpublished doctoral dissertation, Mississippi State University.

Bankston, J. R. (1993). *Instructional leadership behaviors of a selected group of principals in northeast Texas.* Unpublished doctoral dissertation, East Texas State University.

Barling, J., Weber, T., & Kelloway, E. K. (1996). Effects of transformational leadership training on attitudinal and financial outcomes: A field experiment. *Journal of Applied Psychology, 81* (6), 827–832.

Bendo, J. J. (1984). *Application of counseling and communication techniques to supervisory training in industry.* Unpublished doctoral dissertation, University of Akron.

Birkenbach, X. C., Karnfer, L., & ter Morshuizen, J. D. (1985). The development and the evaluation of a behaviour-modelling training programme for supervisors. *South African Journal of Psychology, 15* (1), 11–19.

Briddell, W. B. (1986). *The effects of a time management training program upon occupational stress levels and the Type A behavioral pattern in college administrators.* Unpublished doctoral dissertation, Florida State University.

Bruwelheide, L. R., & Duncan, P. K. (1985). A method for evaluating corporation training seminars. *Journal of Organizational Behavior Management, 7* (1–2), 65–94.

Cato, B. (1990). Effectiveness of a management training institute. *Journal of Park and Recreation Administration, 8* (3), 38–46.

Clark, B. R. (1989). *Professional development emphasizing the socialization progress for academicians planning administrative leadership roles.* Unpublished doctoral dissertation, George Washington University.

Clark, H. B., Wood, R., Kuehnel, T., Flanagan, S., Mosk, M., & Northrup, J. T. (1985). Preliminary validation and training of supervisory interactional skills. *Journal of Organizational Behavior Management, 7* (1–2), 95–115.

Colan, N., & Schneider, R. (1992). The effectiveness of supervisor training: One-year follow-up. *Journal of Employee Assistance Research, 1* (1), 83–95.

Couture, S. (1987). *The effect of an introductory management inservice program on the knowledge of management skills of assistant head nurses at a university teaching hospital.* Unpublished master's thesis, University of Texas Medical Branch at Galveston.

Davis, B. L., & Mount, M. K. (1984). Effectiveness of performance appraisal training using computer assisted instruction and behavior modeling. *Personnel Psychology, 37,* 439–452.

Deci, E. L., Connell, J. P., & Ryan, R. M. (1989). Self-determination in a work organization. *Journal of Applied Psychology, 74* (4), 580–590.

DeNisi, A. S., & Peters, L. H. (1996). Organization of information in memory and the performance appraisal process: Evidence from the field. *Journal of Applied Psychology, 81* (6), 717–737.

DePiano, L. G., & McClure, L. F. (1987). The training of principals for school advisory council leadership. *Journal of Community Psychology, 15,* 253–267.

Devlin-Scherer, W. L., Devlin-Scherer, R., Wright, W., Roger, A. M., & Meyers, K. (1997). The effects of collaborative teacher study groups and principal coaching on individual teacher change. *Journal of Classroom Interaction, 32* (1), 18–22.

Donohoe, T. L., Johnson, J. T., & Stevens, J. (1997). An analysis of an employee assistance supervisory training program. *Employee Assistance Quarterly, 12* (3), 25–34.

Dorfman, P. W., Stephan, W. G., & Loveland, J. (1986). Performance appraisal behaviors: Supervisor perceptions and subordinate reactions. *Personnel Psychology, 39* (3), 579–597.

Dvir, T., Eden, D., Avolio, B. J., & Shamir, B. (2002). Impact of transformational leadership on follower development and performance: A field experiment. *Academy of Management Journal, 45,* 735–744.

Earley, P. C. (1987). Intercultural training for managers: A comparison of documentary and interpersonal methods. *Academy of Management Journal, 30,* 685–698.

Eden, D. (1986). Team development: Quasi-experimental confirmation among combat companies. *Group and Organization Studies, 11* (3), 133–146.

Eden, D., Geller, D., Gewirtz, A., Gordon-Terner, R., Inbar, I., Liberman, M., Pass, Y., Salomon-Segev, I., & Shalit, M. (2000). Implanting Pygmalion leadership style through workshop training: Seven field experiments. *Leadership Quarterly, 11* (2), 171–210.

Edwards, W. S. (1992). *Training secondary school administrators for crisis management.* Unpublished doctoral dissertation, University of Georgia.

Faerman, S. R., & Ban, C. (1993). Trainee satisfaction and training impact: Issues in training evaluation. *Public Productivity and Management Review, 16* (3), 299–314.

Frost, Dean E. (1986). A test of situational engineering for training leaders. *Psychological Reports, 59* (2), 771–782.

Fuller, A. G. (1985). *A study of an organizational development program for leadership effectiveness in higher education.* Unpublished doctoral dissertation. Memphis State University.

Gerstein, L. H., Eichenhofer, D. J., Bayer, G. A., Valutis, W., & Janowski, J. (1989). EAP referral training and supervisors' beliefs about troubled workers. *Employee Assistance Quarterly, 4* (4), 15–30.

Graen, G., Novak, M., & Sommerkamp, P. (1982). The effects of leader-member exchange and job design on productivity and satisfaction: Testing a dual attachment model. *Organizational Behavior and Human Performance, 30,* 109–131.

Haccoun, R. R., & Hamtiaux, T. (1994). Optimizing knowledge tests for inferring learning acquisition levels in single group training evaluation designs: The internal referencing strategy. *Personnel Psychology, 47,* 593–604.

Harrison, J. K. (1992). Individual and combined effects of behavior modeling and the cultural assimilator in cross-cultural management training. *Journal of Applied Psychology,* 77 (6), 952–962.

Henry, J. E. (1983). *The effects of team building, upper management support, and level of manager on the application of management training in performance appraisal skills as perceived by subordinates and self.* Unpublished doctoral dissertation, University of Colorado.

Hill, S. (1992). *Leadership development training for principals: Its initial impact on school culture.* Unpublished doctoral dissertation, Baylor University.

Innami, I. (1994). The quality of group decisions, group behavior, and intervention. *Organizational Behavior and Human Decision Processes, 60,* 409–430.

Ivancevich, J. M. (1982). Subordinates' reactions to performance appraisal interviews: A test of feedback and goal-setting techniques. *Journal of Applied Psychology, 67* (5), 581–587.

Jalbert, N. M., Morrow, C., & Joglekar, A. (2000, April 14–16). *Measuring change with 360s: Using a degree of change measure versus performance.* Paper presented at the Fifteenth Annual Conference of the Society for Industrial and Organizational Psychology, New Orleans.

Josefowitz, A. J. (1984). *The effects of management development training on organizational climate.* Unpublished doctoral dissertation, University of Minnesota.

Katzenmeyer, M. H. (1988). *An evaluation of behavior modeling training designed to improve selected skills of educational managers.* Unpublished doctoral dissertation, Florida State University.

Klein, E. B., Kossek, E. E., & Astrachan, J. H. (1992). Affective reactions to leadership education: An exploration of the same-gender effect. *Journal of Applied Behavioral Science, 28* (1), 102–117.

Lafferty, B. D. (1998). *Investigation of a leadership development program: An empirical investigation of a leadership development program.* Unpublished doctoral dissertation, George Washington University.

Larkin, J. M. (1996). *Effects of a conflict management course on nurse managers.* Unpublished master's thesis, Duquesne University.

Larsen, J. E. (1983). *An evaluative study of the effect of self-management training on group leadership training program for hospital managers.* Unpublished doctoral dissertation, University of Nebraska.

Lawrence, H. V., & Wiswell, A. K. (1993). Using the work group as a laboratory for learning: Increasing leadership and team effectiveness through feedback. *Human Resource Development Quarterly, 4* (2), 135–148.

Martineau, J. W. (1995). *A contextual examination of the effectiveness of a supervisory skills training program.* Unpublished doctoral dissertation, Pennsylvania State University.

Mattox, R. J. (1985). *Self, subordinate, and supervisor reports on the long-term effects of a management training program.* Unpublished doctoral dissertation, East Texas State University.

Maurer, S. D., & Fay, C. (1988). Effect of situational interviews, conventional structured interviews, and training on interview rating agreement: An experimental analysis. *Personnel Psychology, 41* (2), 329–344.

May, G. L., & Kahnweiler, W. M. (2000). The effect of a mastery practice design on learning and transfer in behavior modeling training. *Personnel Psychology, 53* (2), 353–373.

McCauley, C. D., & Hughes-James, M. W. (1994). *An evaluation of the outcomes of a leadership development program.* Greensboro, NC: Center for Creative Leadership.

McCredie, H., & Hool, J. (1992). The validation, learning dynamics and application of an off-the-shelf course in analytical skills. *Management Education and Development, 23* (4), 292–306.

Moxnes, P., & Eilertsen, D. (1991). The influence of management training upon organizational climate: An exploratory study. *Journal of Organizational Behavior, 12,* 399–411.

Nelson, E. P. (1990). *The impact of the Springfield development program on principals' administrative behavior.* Unpublished doctoral dissertation, University of Minnesota.

Niska, J. M. (1991). *Examination of a cooperative learning supervision training and development model.* Unpublished doctoral dissertation, Iowa State University.

Paquet, B., Kasl, E., Weinstein, L., & Waite, W. (1987). The bottom line. *Training and Development Journal, 41* (5), 27–33.

Parker, S. K. (1998). Enhancing role breadth self-efficacy: The roles of job enrichment and other organizational interventions. *Journal of Applied Psychology, 83* (6), 835–852.

Posner, C. S. (1982). *An inferential analysis measuring the cost effectiveness of training outcomes of a supervisory development program using four selected economic indices: An experimental study in a public agency setting.* Unpublished doctoral dissertation, Ball State University.

Reaves, W. M. (1993). *The impact of leadership training on leadership effectiveness.* Unpublished doctoral dissertation, North Carolina State University.

Robertson, M. M. (1992). *Crew resource management training program for maintenance technical operators: Demonstration of a systematic training evaluation model.* Unpublished doctoral dissertation, University of Southern California.

Rosti, R. T., Jr., & Shipper, F. (1998). A study of the impact of training in a management development program based on 360 feedback. *Journal of Managerial Psychology, 13* (1–2), 77–89.

Russell, J. S., Wexley, K. N., & Hunter, J. E. (1984). Questioning the effectiveness of behavior modeling training in an industrial setting. *Personnel Psychology, 37,* 465–481.

Savan, M. (1983). *The effects of leadership training program on supervisory learning and performance.* Unpublished doctoral dissertation, Union for Experimenting Colleges and Universities, Cincinnati, Ohio.

Scandura, T. A., & Graen, G. B. (1984). Moderating effects of initial leader-member exchange status on the effects of a leadership intervention. *Journal of Applied Psychology, 69* (3), 428–436.

Schneider, R. J., & Colan, N. B. (1992). The effectiveness of employee assistance program supervisory training: An experimental study. *Human Resource Development Quarterly 3* (4), 345–356.

Shipper, F., & Neck, C. P. (1990). Subordinates' observations: Feedback for management development. *Human Resource Development Quarterly, 1* (4), 371–385.

Smith, R. M., Montello, P. A., & White, P. E. (1992). Investigation of interpersonal management training for educational administrators. *Journal of Educational Research, 85* (4), 242–245.

Sniderman, R. L. (1992). *The use of SYMLOG in the evaluation of the effectiveness of a management development program.* Unpublished doctoral dissertation, California School for Professional Psychology, Los Angeles.

Sogunro, O. A. (1997). Impact of training on leadership development: Lessons from a leadership training program. *Evaluation Review, 21* (6), 713–737.
Steele, R. W. (1984). *The effect of interactive skills training on middle managers in a major corporation.* Unpublished doctoral dissertation, George Washington University.
Tenorio, E. M. (1996). *A level three evaluation of a management training program.* Unpublished master's thesis, San Jose State University.
Tesoro, F. M. (1991). *The use of the measurement of continuous improvement model for training program evaluation.* Unpublished doctoral dissertation, Purdue University.
Tharenou, P., & Lyndon, J. T. (1990). The effect of a supervisory development program on leadership style. *Journal of Business and Psychology, 4* (3), 365–373.
Thoms, P., & Greenberger, D. B. (1998). A test of vision training and potential antecedents to leaders' visioning ability. *Human Resource Development Quarterly, 9* (1), 3–19.
Thoms, P., & Klein, H. J. (1994). Participation and evaluative outcomes in management training. *Human Resource Development Quarterly, 5* (1), 27–39.
Tracey, J. B., Tannenbaum, S. I., & Kavanaugh, M. J. (1995). Applying trained skills on the job: The importance of the work environment. *Journal of Applied Psychology, 80* (2), 239–252.
Tziner, A., Haccoun, R. R., & Kadish, A. (1991). Personal and situational characteristics influencing the effectiveness of transfer of training improvement strategies. *Journal of Occupational Psychology, 64,* 167–177.
Urban, T. F., Ferris, G. R., Crowe, D. F., & Miller, R. L. (1985, March). Management training: Justify costs or say goodbye. *Training and Development Journal,* 68–71.
Warr, P., & Bunce, D. (1995). Trainee characteristics and the outcomes of open learning. *Personnel Psychology, 48,* 347–375.
Werle, T. (1985). *The development, implementation and evaluation of a training program for middle-managers.* Unpublished doctoral dissertation, Union Graduate School.
Williams, A. J. (1993). *The effects of police officer supervisory training on self-esteem and leadership style adaptability considering the impact of life experiences.* Unpublished doctoral dissertation, Georgia State University.
Wolf, M. S. (1996). Changes in leadership styles as a function of a four-day leadership training institute for nurse managers: A perspective on continuing education program evaluation. *Journal of Continuing Education in Nursing, 27* (6), 245–252.
Woods, E. F. (1987). *The effects of the Center for Educational Administrator Development management training program on site administrators' leadership behavior.* Unpublished doctoral dissertation, University of Southern California.
Yang, Z. Z. (1988). *An exploratory study of the impact of a Western management training program on Chinese managers.* Unpublished doctoral dissertation, Pepperdine University.
Yaworksy, B. (1994). *A study of management training in a criminal justice setting.* Unpublished doctoral dissertation, Rutgers University.
Young, D. P., & Dixon, N. M. (1995). Extending leadership development beyond the classroom: Looking at process and outcomes. In *Proceedings of the Academy of Human Resource Development,* 5–1. Baton Rouge, LA: Academy of Human Resource Management.

Doris B. Collins is associate vice chancellor for student life and academic services at Louisiana State University.

Elwood F. Holton III is the Jones S. Davis Professor of Human Resource, Leadership and Organization Development in the School of Human Resource Education at Louisiana State University.