

# Effectiveness of Training in Organizations: A Meta-Analysis of Design and Evaluation Features

Winfred Arthur Jr.  
Texas A&M University

Winston Bennett Jr.  
Air Force Research Laboratory

Pamela S. Edens and Suzanne T. Bell  
Texas A&M University

The authors used meta-analytic procedures to examine the relationship between specified training design and evaluation features and the effectiveness of training in organizations. Results of the meta-analysis revealed training effectiveness sample-weighted mean *ds* of 0.60 ( $k = 15$ ,  $N = 936$ ) for reaction criteria, 0.63 ( $k = 234$ ,  $N = 15,014$ ) for learning criteria, 0.62 ( $k = 122$ ,  $N = 15,627$ ) for behavioral criteria, and 0.62 ( $k = 26$ ,  $N = 1,748$ ) for results criteria. These results suggest a medium to large effect size for organizational training. In addition, the training method used, the skill or task characteristic trained, and the choice of evaluation criteria were related to the effectiveness of training programs. Limitations of the study along with suggestions for future research are discussed.

The continued need for individual and organizational development can be traced to numerous demands, including maintaining superiority in the marketplace, enhancing employee skills and knowledge, and increasing productivity. Training is one of the most pervasive methods for enhancing the productivity of individuals and communicating organizational goals to new personnel. In 2000, U.S. organizations with 100 or more employees budgeted to spend \$54 billion on formal training ("Industry Report," 2000). Given the importance and potential impact of training on organizations and the costs associated with the development and implementation of training, it is important that both researchers and practitioners have a better understanding of the relationship between design and evaluation features and the effectiveness of training and development efforts.

Meta-analysis quantitatively aggregates the results of primary studies to arrive at an overall conclusion or summary across these studies. In addition, meta-analysis makes it possible to assess relationships not investigated in the original primary studies. These, among others (see Arthur, Bennett, & Huffcutt, 2001), are some of the advantages of meta-analysis over narrative reviews. Although there have been a multitude of meta-analyses in other domains of industrial/organizational psychology (e.g., cognitive ability, employment interviews, assessment centers, and

employment-related personality testing) that now allow researchers to make broad summary statements about observable effects and relationships in these domains, summaries of the training effectiveness literature appear to be limited to the periodic narrative *Annual Reviews*. A notable exception is Burke and Day (1986), who, however, limited their meta-analysis to the effectiveness of only managerial training.

Consequently, the goal of the present article is to address this gap in the training effectiveness literature by conducting a meta-analysis of the relationship between specified design and evaluation features and the effectiveness of training in organizations. We accomplish this goal by first identifying design and evaluation features related to the effectiveness of organizational training programs and interventions, focusing specifically on those features over which practitioners and researchers have a reasonable degree of control. We then discuss our use of meta-analytic procedures to quantify the effect of each feature and conclude with a discussion of the implications of our findings for both practitioners and researchers.

## Overview of Design and Evaluation Features Related to the Effectiveness of Training

Over the past 30 years, there have been six cumulative reviews of the training and development literature (Campbell, 1971; Goldstein, 1980; Latham, 1988; Salas & Cannon-Bowers, 2001; Tannenbaum & Yukl, 1992; Wexley, 1984). On the basis of these and other pertinent literature, we identified several design and evaluation features that are related to the effectiveness of training and development programs. However, the scope of the present article is limited to those features over which trainers and researchers have a reasonable degree of control. Specifically, we focus on (a) the type of evaluation criteria, (b) the implementation of training needs assessment, (c) the skill or task characteristics trained, and

---

Winfred Arthur Jr., Pamela S. Edens, and Suzanne T. Bell, Department of Psychology, Texas A&M University; Winston Bennett Jr., Air Force Research Laboratory, Warfighter Training Research Division, Mesa, Arizona.

This research is based in part on Winston Bennett Jr.'s doctoral dissertation, completed in 1995 at Texas A&M University and directed by Winfred Arthur Jr.

Correspondence concerning this article should be addressed to Winfred Arthur Jr., Department of Psychology, Texas A&M University, College Station, Texas 77843-4235. E-mail: wea@psyc.tamu.edu

(d) the match between the skill or task characteristics and the training delivery method. We consider these to be factors that researchers and practitioners could manipulate in the design, implementation, and evaluation of organizational training programs.

### *Training Evaluation Criteria*

The choice of evaluation criteria (i.e., the dependent measure used to operationalize the effectiveness of training) is a primary decision that must be made when evaluating the effectiveness of training. Although newer approaches to, and models of, training evaluation have been proposed (e.g., Day, Arthur, & Gettman, 2001; Kraiger, Ford, & Salas, 1993), Kirkpatrick's (1959, 1976, 1996) four-level model of training evaluation and criteria continues to be the most popular (Salas & Canon-Bowers, 2001; Van Buren & Erskine, 2002). We used this framework because it is conceptually the most appropriate for our purposes. Specifically, within the framework of Kirkpatrick's model, questions about the effectiveness of training or instruction programs are usually followed by asking, "Effective in terms of what? Reactions, learning, behavior, or results?" Thus, the objectives of training determine the most appropriate criteria for assessing the effectiveness of training.

*Reaction criteria*, which are operationalized by using self-report measures, represent trainees' affective and attitudinal responses to the training program. However, there is very little reason to believe that how trainees feel about or whether they like a training program tells researchers much, if anything, about (a) how much they learned from the program (*learning criteria*), (b) changes in their job-related behaviors or performance (*behavioral criteria*), or (c) the utility of the program to the organization (*results criteria*). This is supported by the lack of relationship between reaction criteria and the other three criteria (e.g., Alliger & Janak, 1989; Alliger, Tannenbaum, Bennett, Traver, & Shotland, 1997; Arthur, Tubre, Paul, & Edens, 2003; Colquitt, LePine, & Noe, 2000; Kaplan & Pascoe, 1977; Noe & Schmitt, 1986). In spite of the fact that "reaction measures are not a suitable surrogate for other indexes of training effectiveness" (Tannenbaum & Yukl, 1992, p. 425), anecdotal and other evidence suggests that reaction measures are the most widely used evaluation criteria in applied settings. For instance, in the American Society of Training and Development 2002 *State-of-the-Industry Report*, 78% of the benchmarking organizations surveyed reported using reaction measures, compared with 32%, 9%, and 7% for learning, behavioral, and results, respectively (Van Buren & Erskine, 2002).

*Learning criteria* are measures of the learning outcomes of training; they are not measures of job performance. They are typically operationalized by using paper-and-pencil and performance tests. According to Tannenbaum and Yukl (1992), "trainee learning appears to be a necessary but not sufficient prerequisite for behavior change" (p. 425). In contrast, *behavioral criteria* are measures of actual on-the-job performance and can be used to identify the effects of training on actual work performance. Issues pertaining to the transfer of training are also relevant here. Behavioral criteria are typically operationalized by using supervisor ratings or objective indicators of performance. Although learning and behavioral criteria are conceptually linked, researchers have had limited success in empirically demonstrating this relationship (Alliger et al., 1997; Severin, 1952; cf. Colquitt et al., 2000). This

is because behavioral criteria are susceptible to environmental variables that can influence the transfer or use of trained skills or capabilities on the job (Arthur, Bennett, Stanush, & McNelly, 1998; Fecteau, Dobbins, Russell, Ladd, & Kudisch, 1995; Quiñones, 1997; Quiñones, Ford, Sego, & Smith, 1995; Tracey, Tannenbaum, & Kavanagh, 1995). For example, the posttraining environment may not provide opportunities for the learned material or skills to be applied or performed (Ford, Quiñones, Sego, & Speer-Sorra, 1992). Finally, *results criteria* (e.g., productivity, company profits) are the most distal and macro criteria used to evaluate the effectiveness of training. Results criteria are frequently operationalized by using utility analysis estimates (Cascio, 1991, 1998). Utility analysis provides a methodology to assess the dollar value gained by engaging in specified personnel interventions including training.

In summary, it is our contention that given their characteristic feature of capturing different facets of the criterion space—as illustrated by their weak intercorrelations reported by Alliger et al. (1997)—the effectiveness of a training program may vary as a function of the criteria chosen to measure effectiveness (Arthur, Tubre, et al., 2003). Thus, it is reasonable to ask whether the effectiveness of training—operationalized as effect size *ds*—varies systematically as a function of the outcome criterion measure used. For instance, all things being equal, are larger effect sizes obtained for training programs that are evaluated by using learning versus behavioral criteria? It is important to clarify that criterion type is not an independent or causal variable in this study. Our objective is to investigate whether the operationalization of the dependent variable is related to the observed training outcomes (i.e., effectiveness). Thus, the evaluation criteria (i.e., reaction, learning, behavioral, and results) are simply different operationalizations of the effectiveness of training. Consequently, our first research question is this: Are there differences in the effectiveness of training (i.e., the magnitude of the *ds*) as a function of the operationalization of the dependent variable?

### *Conducting a Training Needs Assessment*

*Needs assessment*, or *needs analysis*, is the process of determining the organization's training needs and seeks to answer the question of whether the organization's needs, objectives, and problems can be met or addressed by training. Within this context, needs assessment is a three-step process that consists of organizational analysis (e.g., Which organizational goals can be attained through personnel training? Where is training needed in the organization?), task analysis (e.g., What must the trainee learn in order to perform the job effectively? What will training cover?), and person analysis (e.g., Which individuals need training and for what?).

Thus, conducting a systematic needs assessment is a crucial initial step to training design and development and can substantially influence the overall effectiveness of training programs (Goldstein & Ford, 2002; McGehee & Thayer, 1961; Sleezer, 1993; Zemke, 1994). Specifically, a systematic needs assessment can guide and serve as the basis for the design, development, delivery, and evaluation of the training program; it can be used to specify a number of key features for the implementation (input) and evaluation (outcomes) of training programs. Consequently, the presence and comprehensiveness of a needs assessment should be

related to the overall effectiveness of training because it provides the mechanism whereby the questions central to successful training programs can be answered. In the design and development of training programs, systematic attempts to assess the training needs of the organization, identify the job requirements to be trained, and identify who needs training and the kind of training to be delivered should result in more effective training. Thus, the research objective here was to determine the relationship between needs assessment and training outcomes.

### *Match Between Skills or Tasks and Training Delivery Methods*

A product of the needs assessment is the specification of the training objectives that, in turn, identifies or specifies the skills and tasks to be trained. A number of typologies have been offered for categorizing skills and tasks (e.g., Gagne, Briggs, & Wagner, 1992; Rasmussen, 1986; Schneider & Shiffrin, 1977). Given the fair amount of overlap between them, they can all be summarized into a general typology that classifies both skills and tasks into three broad categories: cognitive, interpersonal, and psychomotor (Farina & Wheaton, 1973; Fleishman & Quaintance, 1984; Goldstein & Ford, 2002). Cognitive skills and tasks are related to the thinking, idea generation, understanding, problem solving, or the knowledge requirements of the job. Interpersonal skills and tasks are those that are related to interacting with others in a workgroup or with clients and customers. They entail a wide variety of skills including leadership skills, communication skills, conflict management skills, and team-building skills. Finally, psychomotor skills involve the use of the musculoskeletal system to perform behavioral activities associated with a job. Thus, psychomotor tasks are physical or manual activities that involve a range of movement from very fine to gross motor coordination.

Practitioners and researchers have limited control over the choice of skills and tasks to be trained because they are primarily specified by the job and the results of the needs assessment and training objectives. However, they have more latitude in the choice and design of the training delivery method and the match between the skill or task and the training method. For a specific task or training content domain, a given training method may be more effective than others. Because all training methods are capable of, and indeed are intended to, communicate specific skill, knowledge, attitudinal, or task information to trainees, different training methods can be selected to deliver different content (i.e., skill, knowledge, attitudinal, or task) information. Thus, the effect of skill or task type on the effectiveness of training is a function of the match between the training delivery method and the skill or task to be trained. Wexley and Latham (2002) highlighted the need to consider skill and task characteristics in determining the most effective training method. However, there has been very little, if any, primary research directly assessing these effects. Thus, the research objective here was to assess the effectiveness of training as a function of the skill or task trained and the training method used.

### *Research Questions*

On the basis of the issues raised in the preceding sections, this study addressed the following questions:

1. Does the effectiveness of training—operationalized as effect

size *ds*—vary systematically as a function of the evaluation criteria used? For instance, because the effect of extratraining constraints and situational factors increases as one moves from learning to results criteria, will the magnitude of observed effect sizes decrease from learning to results criteria?

2. What is the relationship between needs assessment and training effectiveness? Specifically, will studies with more comprehensive needs assessments be more effective (i.e., obtain larger effect sizes) than those with less comprehensive needs assessments?

3. What is the observed effectiveness of specified training methods as a function of the skill or task being trained? It should be noted that because we expected effectiveness to vary as a function of the evaluation criteria used, we broke down all moderators by criterion type.

## **Method**

### *Literature Search*

For the present study, we reviewed the published training and development literature from 1960 to 2000. We considered the period post-1960 to be characterized by increased technological sophistication in training design and methodology and by the use of more comprehensive training evaluation techniques and statistical approaches. The increased focus on quantitative methods for the measurement of training effectiveness is critical for a quantitative review such as this study. Similar to past training and development reviews (e.g., Latham, 1988; Tannenbaum & Yukl, 1992; Wexley, 1984), the present study also included the practitioner-oriented literature if those studies met the criteria for inclusion as outlined below. Therefore, the literature search encompassed studies published in journals, books or book chapters, conference papers and presentations, and dissertations and theses that were related to the evaluation of an organizational training program or those that measured some aspect of the effectiveness of organizational training.

An extensive literature search was conducted to identify empirical studies that involved an evaluation of a training program or measured some aspects of the effectiveness of training. This search process started with a search of nine computer databases (Defense Technical Information Center, Econlit, Educational Research Information Center, Government Printing Office, National Technical Information Service, PsycLIT/PsycINFO, Social Citations Index, Sociofile, and Wilson) using the following key words: *training effectiveness*, *training evaluation*, *training efficiency*, and *training transfer*. The electronic search was supplemented with a manual search of the reference lists from past reviews of the training literature (e.g., Alliger et al., 1997; Campbell, 1971; Goldstein, 1980; Latham, 1988; Tannenbaum & Yukl, 1992; Wexley, 1984). A review of the abstracts obtained as a result of this initial search for appropriate content (i.e., empirical studies that actually evaluated an organizational training program or measured some aspect of the effectiveness of organizational training), along with a decision to retain only English language articles, resulted in an initial list of 383 articles and papers. Next, the reference lists of these sources were reviewed. As a result of these efforts, an additional 253 sources were identified, resulting in a total preliminary list of 636 sources. Each of these was then reviewed and considered for inclusion in the meta-analysis.

### *Inclusion Criteria*

A number of decision rules were used to determine which studies would be included in the meta-analysis. First, to be included in the meta-analysis, a study must have investigated the effectiveness of an organizational training program or have conducted an empirical evaluation of an organizational training method or approach. Studies evaluating the effectiveness of other training programs were excluded because such programs were

considered to be qualitatively different from more traditional organizational training studies or programs. Second, to be included, studies had to report sample sizes along with other pertinent information. This information included statistics that allowed for the computation of a *d* statistic (e.g., group means and standard deviations). If studies reported statistics such as correlations, univariate *F*, *t*,  $\chi^2$ , or some other test statistic, these were converted to *ds* by using the appropriate conversion formulas (see Arthur et al., 2001, Appendix C, for a summary of conversion formulas). Finally, studies based on single group pretest–posttest designs were excluded from the data.

### Data Set

**Nonindependence.** As a result of the inclusion criteria, an initial data set of 1,152 data points (*ds*) from 165 sources was obtained. However, some of the data points were nonindependent. Multiple effect sizes or data points are nonindependent if they are computed from data collected from the same sample of participants. Decisions about nonindependence also have to take into account whether or not the effect sizes represent the same variable or construct (Arthur et al., 2001). For instance, because criterion type was a variable of interest, if a study reported effect sizes for multiple criterion types (e.g., reaction and learning), these effect sizes were considered to be independent even though they were based on the same sample; therefore, they were retained as separate data points. Consistent with this, data points based on multiple measures of the same criterion (e.g., reactions) for the same sample were considered to be nonindependent and were subsequently averaged to form a single data point. Likewise, data points based on temporally repeated measures of the same or similar criterion for the same sample were also considered to be nonindependent and were subsequently averaged to form a single data point. The associated time intervals were also averaged. Implementing these decision rules resulted in 405 independent data points from 164 sources.

**Outliers.** We computed Hufcutt and Arthur's (1995) and Arthur et al.'s (2001) sample-adjusted meta-analytic deviancy statistic to detect outliers. On the basis of these analyses, we identified 8 outliers. A detailed review of these studies indicated that they displayed unusual characteristics such as extremely large *ds* (e.g., 5.25) and sample sizes (e.g., 7,532). They were subsequently dropped from the data set. This resulted in a final data set of 397 independent *ds* from 162 sources. Three hundred ninety-three of the data points were from journal articles, 2 were from conference papers, and 1 each were from a dissertation and a book chapter. A reference list of sources included in the meta-analysis is available from Winfred Arthur, Jr., upon request.

### Description of Variables

This section presents a description of the variables that were coded for the meta-analysis.

**Evaluation criteria.** Kirkpatrick's (1959, 1976, 1996) evaluation criteria (i.e., reaction, learning, behavioral, and results) were coded. Thus, for each study, the criterion type used as the dependent variable was identified. The interval (i.e., number of days) between the end of training and collection of the criterion data was also coded.

**Needs assessment.** The needs assessment components (i.e., organization, task, and person analysis) conducted and reported in each study as part of the training program were coded. Consistent with our decision to focus on features over which practitioners and researchers have a reasonable degree of control, a convincing argument can be made that most training professionals have some latitude in deciding whether to conduct a needs assessment and its level of comprehensiveness. However, it is conceivable, and may even be likely, that some researchers conducted a needs assessment but did not report doing so in their papers or published works. On the other hand, we also think that it can be reasonably argued that if there was no mention of a needs assessment, then it was probably not

a very potent study variable or manipulation. Consequently, if a study did not mention conducting a needs assessment, this variable was coded as "missing." We recognize that this may present a weak test of this research question, so our analyses and the discussion of our results are limited to only those studies that reported conducting some needs assessment.

**Training method.** The specific methods used to deliver training in the study were coded. Multiple training methods (e.g., lectures and discussion) were recorded if they were used in the study. Thus, the data reflect the effect for training methods as reported, whether single (e.g., audiovisual) or multiple (e.g., audiovisual and lecture) methods were used.

**Skill or task characteristics.** Three types of training content (i.e., cognitive, interpersonal, and psychomotor) were coded. For example, if the focus of the training program was to train psychomotor skills and tasks, then the psychomotor characteristic was coded as 1 whereas the other characteristics (i.e., cognitive and interpersonal) were coded as 0. Skill or task types were generally nonoverlapping—only 14 (4%) of the 397 data points in the final data set focused on more than one skill or task.

### Coding Accuracy and Interrater Agreement

The coding training process and implementation were as follows. First, Winston Bennett, Jr., and Pamela S. Edens were furnished with a copy of a coder training manual and reference guide that had been developed by Winfred Arthur, Jr., and Winston Bennett, Jr., and used with other meta-analysis projects (e.g., Arthur et al., 1998; Arthur, Day, McNelly, & Edens, in press). Each coder used the manual and reference guide to independently code 1 article. Next, they attended a follow-up training meeting with Winfred Arthur, Jr., to discuss problems encountered in using the guide and the coding sheet and to make changes to the guide or the coding sheet as deemed necessary. They were then assigned the same 5 articles to code. After coding these 5 articles, the coders attended a second training session in which the degree of convergence between them was assessed. Discrepancies and disagreements related to the coding of the 5 articles were resolved by using a consensus discussion and agreement among the authors. After this second meeting, Pamela S. Edens subsequently coded the articles used in the meta-analysis. As part of this process, Winston Bennett, Jr., coded a common set of 20 articles that were used to assess the degree of interrater agreement. The level of agreement was generally high, with a mean overall agreement of 92.80% ( $SD = 5.71$ ).

### Calculating the Effect Size Statistic (*d*) and Analyses

**Preliminary analyses.** The present study used the *d* statistic as the common effect-size metric. Two hundred twenty-one (56%) of the 397 data points were computed by using means and standard deviations presented in the primary studies. The remaining 176 data points (44%) were computed from test statistics (i.e., correlations, *t* statistics, or univariate two-group *F* statistics) that were converted to *ds* by using the appropriate conversion formulas (Arthur et al., 2001; Dunlap, Cortina, Vaslow, & Burke, 1996; Glass, McGaw, & Smith, 1981; Hunter & Schmidt, 1990; Wolf, 1986).

The data analyses were performed by using Arthur et al.'s (2001) SAS PROC MEANS meta-analysis program to compute sample-weighted means. Sample weighting assigns studies with larger sample sizes more weight and reduces the effect of sampling error because sampling error generally decreases as the sample size increases (Hunter & Schmidt, 1990). We also computed 95% confidence intervals (CIs) for the sample-weighted mean *ds*. CIs are used to assess the accuracy of the estimate of the mean effect size (Whitener, 1990). CIs estimate the extent to which sampling error remains in the sample-size-weighted mean effect size. Thus, CI gives the range of values that the mean effect size is likely to fall within if other sets of studies were taken from the population and used in the meta-analysis. A desirable CI is one that does not include zero if a nonzero relationship is hypothesized.

**Moderator analyses.** In the meta-analysis of effect sizes, the presence of one or more moderator variables is suspected when sufficient variance



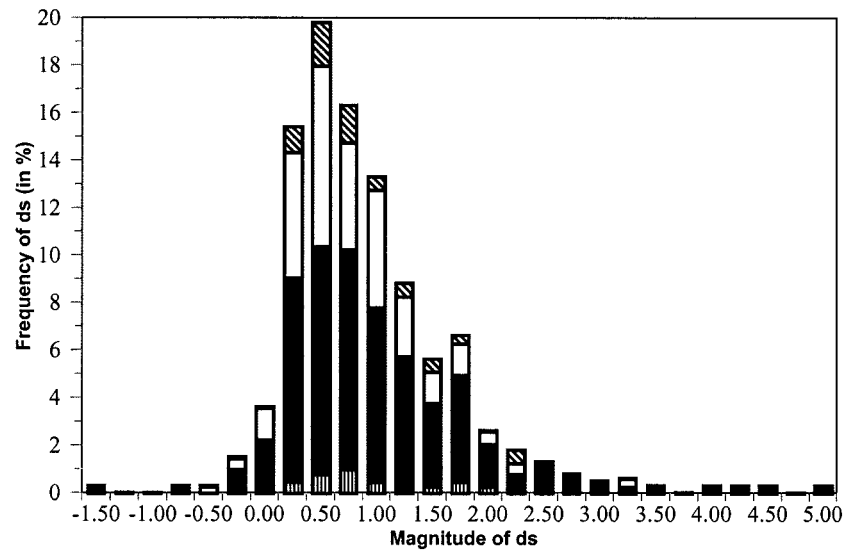


Figure 1. Distribution (histogram) of the 397 *ds* (by criteria) of the effectiveness of organizational training included in the meta-analysis. Values on the *x*-axis represent the upper value of a 0.25 band. Thus, for instance, the value 0.00 represents *ds* falling between  $-0.25$  and  $0.00$ , and  $5.00$  represents *ds* falling between  $4.75$  and  $5.00$ . Diagonal bars indicate results criteria, white bars indicate behavioral criteria, black bars indicate learning criteria, and vertical bars indicate reaction criteria.

remains in the corrected effect size. Alternately, various moderator variables may be suggested by theory. Thus, the decision to search or test for moderators may be either theoretically or empirically driven. In the present study, decisions to test for the presence and effects of moderators were theoretically based. To assess the relationship between each feature and the effectiveness of training, studies were categorized into separate subsets according to the specified level of the feature. An overall, as well as a subset, mean effect size and associated meta-analytic statistics were then calculated for each level of the feature.

For the moderator analysis, the meta-analysis was limited to factors with 2 or more data points. Although there is no magical cutoff as to the minimum number of studies to include in a meta-analysis, we acknowledge that using such a small number raises the possibility of second-order sampling error and concerns about the stability and interpretability of the obtained meta-analytic estimates (Arthur et al., 2001; Hunter & Schmidt, 1990). However, we chose to use such a low cutoff for the sake of completeness but emphasize that meta-analytic effect sizes based on less than 5 data points should be interpreted with caution.

## Results

### Evaluation Criteria

Our first objective for the present meta-analysis was to assess whether the effectiveness of training varied systematically as a function of the evaluation criteria used. Figure 1 presents the distribution (histogram) of *ds* included in the meta-analysis. The *ds* in the figure are grouped in 0.25 intervals. The histogram shows that most of the *ds* were positive, only 5.8% (3.8% learning and 2.0% behavioral criteria) were less than zero. Table 1 presents the results of the meta-analysis and shows the sample-weighted mean *d*, with its associated corrected standard deviation. The corrected standard deviation provides an index of the variation of *ds* across the studies in the data set. The percentage of variance accounted for by sampling error and 95% CIs are also provided.

Consistent with the histogram, the results presented in Table 1 show medium to large sample-weighted mean effect sizes ( $d = 0.60$ – $0.63$ ) for organizational training effectiveness for the four evaluation criteria. (Cohen, 1992, describes *ds* of 0.20, 0.50, and 0.80 as small, medium, and large effect sizes, respectively.)

The largest effect was obtained for learning criteria. However, the magnitude of the differences between criterion types was small; *ds* ranged from 0.63 for learning criteria to 0.60 for reaction criteria. To further explore the effect of criterion type, we limited our analysis to a within-study approach.<sup>1</sup> Specifically, we identified studies that reported multiple criterion measures and assessed the differences between the criterion types. Five sets of studies were available in the data set—those that reported using (a) reaction and learning, (b) learning and behavioral, (c) learning and results, (d) behavioral and results, and (e) learning, behavioral and results criteria. For all comparisons of learning with subsequent criteria (i.e., behavioral and results [with the exception of the learning and results analysis, which was based on 3 data points]), a clear trend that can be garnered from these results is that, consistent with issues of transfer, lack of opportunity to perform, and skill loss, there was a decrease in effect sizes from learning to these criteria. For instance, the average decrease in effect sizes for the learning and behavioral comparisons was 0.77, a fairly large decrease.

Arising from an interest to describe the methodological state of, and publication patterns in, the extant organizational training effectiveness literature, the results presented in Table 1 show that the smallest number of data points were obtained for reaction criteria ( $k = 15$  [4%]). In contrast, the American Society for Training and Development 2002 *State-of-the-Industry Report* (Van Buren &

<sup>1</sup> We thank anonymous reviewers for suggesting this analysis.

Table 1  
*Meta-Analysis Results of the Relationship Between Design and Evaluation Features and the Effectiveness of Organizational Training*

Training design and evaluation features	No. of data points ( <i>k</i> )	<i>N</i>	Sample-weighted <i>M d</i>	Corrected <i>SD</i>	% Variance due to sampling error	95% CI	
						L	U
Evaluation criteria <sup>a</sup>							
Reaction	15	936	0.60	0.26	50.69	0.09	1.11
Learning	234	15,014	0.63	0.59	16.25	−0.53	1.79
Behavioral	122	15,627	0.62	0.29	28.34	0.05	1.19
Results	26	1,748	0.62	0.46	23.36	−0.28	1.52
Multiple criteria within study							
Reaction & learning							
Reaction	12	790	0.59	0.26	49.44	0.08	1.09
Learning	12	790	0.61	0.43	25.87	−0.24	1.46
Learning & behavioral							
Learning	17	839	0.66	0.68	16.32	−0.67	1.98
Behavioral	17	839	0.44	0.50	25.42	−0.55	1.43
Learning & results							
Learning	3	187	0.73	0.59	16.66	−0.44	1.89
Results	3	187	1.42	0.61	18.45	0.23	2.60
Behavioral & results							
Behavioral	10	736	0.91	0.53	17.75	−0.14	1.96
Results	10	736	0.57	0.39	27.48	−0.20	1.33
Learning, behavioral, & results							
Learning	5	258	2.20	0.00	100.00	2.20	2.20
Behavioral	5	258	0.63	0.51	24.54	−0.37	1.63
Results	5	258	0.82	0.13	82.96	0.56	1.08
Needs assessment level							
Reaction							
Organizational only	2	115	0.28	0.00	100.00	0.28	0.28
Learning							
Organizational & person	2	58	1.93	0.00	100.00	1.93	1.93
Organizational only	4	154	1.09	0.84	15.19	−0.55	2.74
Task only	4	230	0.90	0.50	24.10	−0.08	1.88
Behavioral							
Task only	4	211	0.63	0.02	99.60	0.60	0.67
Organizational, task, & person	2	65	0.43	0.00	100.00	0.43	0.43
Organizational only	4	176	0.35	0.00	100.00	0.35	0.35
Skill or task characteristic							
Reaction							
Psychomotor	2	161	0.66	0.00	100.00	0.66	0.66
Cognitive	12	714	0.61	0.31	42.42	0.00	1.23
Learning							
Cognitive & interpersonal	3	106	2.08	0.26	73.72	1.58	2.59
Psychomotor	22	937	0.80	0.31	52.76	0.20	1.41
Interpersonal	65	3,470	0.68	0.69	14.76	−0.68	2.03
Cognitive	143	10,445	0.58	0.55	16.14	−0.50	1.66
Behavioral							
Cognitive & interpersonal	4	39	0.75	0.30	56.71	0.17	1.34
Psychomotor	24	1,396	0.71	0.30	46.37	0.13	1.30
Cognitive	37	11,369	0.61	0.21	23.27	0.20	1.03
Interpersonal	56	2,616	0.54	0.41	35.17	−0.27	1.36
Results							
Interpersonal	7	299	0.88	0.00	100.00	0.88	0.88
Cognitive	9	733	0.60	0.60	12.74	−0.58	1.78
Cognitive & psychomotor	4	292	0.44	0.00	100.00	0.44	0.44
Psychomotor	4	334	0.43	0.00	100.00	0.43	0.43

(table continues)

Training design and evaluation features	No. of data points ( <i>k</i> )	<i>N</i>	Sample- weighted <i>M d</i>	Corrected <i>SD</i>	% Variance due to sampling error	95% CI	
						L	U
Skill or task characteristic by training method: Cognitive skills or tasks							
Reaction							
Self-instruction	2	96	0.91	0.22	66.10	0.47	1.34
C-A instruction	3	246	0.31	0.00	100.00	0.31	0.31
Learning							
Audiovisual & self-instruction	2	46	1.56	0.51	48.60	0.55	2.57
Audiovisual & job aid	2	117	1.49	0.00	100.00	1.49	1.49
Lecture & audiovisual	2	60	1.46	0.00	100.00	1.46	1.46
Lecture, audiovisual, & discussion	5	142	1.35	1.04	14.82	−0.68	3.39
Lecture, self-instruction, & programmed instruction	2	79	1.15	0.00	100.00	1.15	1.15
Audiovisual	9	302	1.06	1.21	8.97	−1.32	3.43
Equipment simulators	10	192	0.87	0.00	100.00	0.87	0.87
Audiovisual & programmed instruction	3	64	0.72	0.08	97.22	0.57	0.88
Audiovisual & C-A instruction	3	363	0.66	0.00	100.00	0.66	0.66
Programmed instruction	15	2,312	0.65	0.35	18.41	−0.04	1.33
Lecture, audiovisual, discussion, C-A instruction, & self-instruction	4	835	0.62	0.56	6.10	−0.48	1.71
Self-instruction	4	162	0.53	0.33	49.53	−0.12	1.18
Lecture & discussion	27	1,518	0.50	0.69	13.84	−0.85	1.85
Lecture	12	1,176	0.45	0.51	14.18	−0.54	1.45
C-A instruction	7	535	0.40	0.06	94.40	0.28	0.51
C-A instruction & self- instruction	4	1,019	0.38	0.46	7.17	−0.51	1.28
C-A instruction & programmed instruction	4	64	0.34	0.00	100.00	0.34	0.34
Discussion	8	427	0.20	0.79	11.10	−1.35	1.75
Behavioral							
Lecture	10	8,131	0.71	0.11	29.00	0.49	0.93
Lecture & audiovisual	3	240	0.66	0.15	70.98	0.37	0.96
Lecture & discussion	6	321	0.43	0.08	93.21	0.28	0.58
Discussion	2	245	0.36	0.00	100.00	0.36	0.36
Lecture, audiovisual, discussion, C-A instruction, & self-instruction	4	640	0.32	0.00	100.00	0.32	0.32
Results							
Lecture & discussion	4	274	0.54	0.41	27.17	−0.26	1.34
Cognitive & interpersonal skills or tasks							
Learning							
Lecture & discussion	2	78	2.07	0.17	48.73	1.25	2.89
Behavioral							
Lecture & discussion	3	128	0.54	0.00	100.00	0.54	0.54
Cognitive & psychomotor skills or tasks							
Results							
Lecture & discussion	2	90	0.51	0.00	100.00	0.51	0.51
Job rotation, lecture, & audiovisual	2	112	0.32	0.00	100.00	0.32	0.32
Interpersonal skills or tasks							
Learning							
Audiovisual	6	247	1.44	0.64	23.79	0.18	2.70
Lecture, audiovisual, & teleconference	2	70	1.29	0.49	37.52	0.32	2.26
Lecture	7	162	0.89	0.51	44.90	−0.10	1.88
Lecture, audiovisual, & discussion	5	198	0.71	0.67	19.92	−0.61	2.04

Training design and evaluation features	No. of data points ( <i>k</i> )	<i>N</i>	Sample- weighted <i>M d</i>	Corrected <i>SD</i>	% Variance due to sampling error	95% CI	
						L	U
Interpersonal skills or tasks ( <i>continued</i> )							
Lecture & discussion	21	1,308	0.70	0.73	11.53	−0.74	2.14
Discussion	14	637	0.61	0.81	12.75	−0.98	2.20
Lecture & audiovisual	4	131	0.34	0.00	100.00	0.34	0.34
Audiovisual & discussion	2	562	0.31	0.00	100.00	0.31	0.31
Behavioral							
Programmed instruction	3	145	0.94	0.00	100.00	0.94	0.94
Audiovisual, programmed instruction, & discussion	4	144	0.75	0.14	86.22	0.47	1.02
Lecture & discussion	21	589	0.64	0.74	22.78	−0.81	2.10
Discussion	6	404	0.56	0.28	44.91	0.01	1.11
Lecture	6	402	0.56	0.00	100.00	0.56	0.56
Lecture, audiovisual, discussion, & self-instruction	2	116	0.44	0.00	100.00	0.44	0.44
Lecture, audiovisual, & discussion	10	480	0.22	0.17	67.12	−0.12	0.56
Results							
Lecture & discussion	3	168	0.79	0.00	100.00	0.79	0.79
Lecture, audiovisual, & discussion	2	51	0.78	0.17	85.74	0.44	1.12
Psychomotor skills or tasks							
Learning							
Audiovisual & discussion	3	156	1.11	0.58	21.58	−0.03	2.24
Lecture & audiovisual	3	242	0.69	0.29	38.63	0.12	1.27
C-A instruction	2	70	0.67	0.00	100.00	0.67	0.67
Behavioral							
Equipment simulators	2	32	1.81	0.00	100.00	1.81	1.81
Audiovisual	3	96	1.45	0.00	100.00	1.45	1.45
Lecture	4	256	0.91	0.31	42.68	0.31	1.52
Lecture, audiovisual, & teleconference	2	56	0.88	0.00	100.00	0.88	0.88
Discussion	4	324	0.67	0.00	100.00	0.67	0.67
Lecture, discussion, & equipment simulators	3	294	0.42	0.00	100.00	0.42	0.42
Results							
Lecture, discussion, & equipment simulators	3	294	0.38	0.00	100.00	0.38	0.38

Note. L = lower; U = upper; C-A = computer-assisted.

<sup>a</sup> The overall sample-weighted mean *d* across the four evaluation criteria was 0.62 (*k* = 397, *N* = 33,325, corrected *SD* = 0.46, % variance due to sampling error = 19.64, 95% confidence interval = -0.27-1.52). It is important to note that because the four evaluation criteria are argued to be conceptually distinct and focus on different facets of the criterion space, there is some question about the appropriateness of an "overall effectiveness" effect size. Accordingly, this information is presented here for the sake of completeness.

Erskine, 2002) indicated that 78% of the organizations surveyed used reaction measures, compared with 32%, 19%, and 7% for learning, behavioral, and results, respectively. The wider use of reaction measures in practice may be due primarily to their ease of collection and proximal nature. This discrepancy between the frequency of use in practice and the published research literature is not surprising given that academic and other scholarly and empirical journals are unlikely to publish a training evaluation study that focuses only or primarily on reaction measures as the criterion for evaluating the effectiveness of organizational training. In further support of this, Alliger et al. (1997), who did not limit their inclusion criteria to organizational training programs like we did, reported only 25 data points for reaction criteria, compared with 89 for learning and 58 for behavioral criteria.

Like reaction criteria, an equally small number of data points were obtained for results criteria (*k* = 26 [7%]). This is probably a function of the distal nature of these criteria, the practical logistical constraints associated with conducting results-level evaluations, and the increased difficulty in controlling for confounding variables such as the business climate. Finally, substantially more data points were obtained for learning and behavioral criteria (*k* = 234 [59%] and 122 [31%], respectively).

We also computed descriptive statistics for the time intervals for the collection of the four evaluation criteria to empirically describe the temporal nature of these criteria. The correlation between interval and criterion type was .41 (*p* < .001). Reaction criteria were always collected immediately after training (*M* = 0.00 days, *SD* = 0.00), followed by learning criteria, which on average, were



collected 26.34 days after training ( $SD = 87.99$ ). Behavioral criteria were more distal ( $M = 133.59$  days,  $SD = 142.24$ ), and results criteria were the most temporally distal ( $M = 158.88$  days,  $SD = 187.36$ ). We further explored this relationship by computing the correlation between the evaluation criterion time interval and the observed  $d$  ( $r = .03$ ,  $p = .56$ ). In summary, these results indicate that although the evaluation criteria differed in terms of the time interval in which they were collected, time intervals were not related to the observed effect sizes.

### *Needs Assessment*

The next research question focused on the relationship between needs assessment and training effectiveness. These analyses were limited to only studies that reported conducting a needs assessment. For these and all other moderator analyses, multiple levels within each factor were ranked in descending order of the magnitude of sample-weighted  $ds$ . We failed to identify a clear pattern of results. For instance, a comparison of studies that conducted only an organizational analysis to those that performed only a task analysis showed that for learning criteria, studies that conducted only an organizational analysis obtained larger effect sizes than those that conducted a task analysis. However, these results were reversed for the behavioral criteria. Furthermore, contrary to what may have been expected, studies that implemented multiple needs assessment components did not necessarily obtain larger effect sizes. On a cautionary note, we acknowledge that these analyses were all based on 4 or fewer data points and thus should be cautiously interpreted. Related to this, it is worth noting that studies reporting a needs assessment represented a very small percentage—only 6% (22 of 397)—of the data points in the meta-analysis.

### *Match Between Skill or Task Characteristics and Training Delivery Method*

Testing for the effect of skill or task characteristics was intended to shed light on the “trainability” of skills and tasks. For both learning and behavioral criteria, the largest effects were obtained for training that included both cognitive and interpersonal skills or tasks (mean  $ds = 2.08$  and  $0.75$ , respectively), followed by psychomotor skills or tasks (mean  $ds = 0.80$  and  $0.71$ , respectively). Medium effects were obtained for both interpersonal and cognitive skills or tasks, although their rank order was reversed for learning and behavioral criteria. Where results criteria were used, the largest effect was obtained for interpersonal skills or tasks (mean  $d = 0.88$ ) and the smallest for psychomotor skills or tasks (mean  $d = 0.43$ ). A medium to large effect was obtained for cognitive skills or tasks. Finally, the data for reaction criterion were limited. Specifically, there were only two skill or task types, psychomotor and cognitive, and for the former, there were only 2 data points. Nevertheless, a medium to large effect size was obtained for both skills or tasks, and unlike results for the other three criterion types, the differences for reaction criteria were minimal.

We next investigated the effectiveness of specified training delivery methods as a function of the skill or task being trained. Again, these data were analyzed by criterion type. The results presented in Table 1 show that very few studies used a single training method. They also indicate a wide range in the mean  $ds$

both within and across evaluation criteria. For instance, for cognitive skills or tasks that used learning criteria, the sample-weighted mean  $ds$  ranged from 1.56 to 0.20. However, overall, the magnitude of the effect sizes was generally favorable and ranged from medium to large. As an example of this, it is worth noting that in contrast to its anecdotal reputation as a boring training method and the subsequent perception of ineffectiveness, the mean  $d$  for lectures (either by themselves or in conjunction with other training methods) were generally favorable across skill or task types and evaluation criteria. In summary, our results suggest that organizational training is generally effective. Furthermore, they also suggest that the effectiveness of training appears to vary as function of the specified training delivery method, the skill or task being trained, and the criterion used to operationalize effectiveness.

### *Discussion*

Meta-analytic procedures were applied to the extant published training effectiveness literature to provide a quantitative “population” estimate of the effectiveness of training and also to investigate the relationship between the observed effectiveness of organizational training and specified training design and evaluation features. Depending on the criterion type, the sample-weighted effect size for organizational training was 0.60 to 0.63, a medium to large effect (Cohen, 1992). This is an encouraging finding, given the pervasiveness and importance of training to organizations. Indeed, the magnitude of this effect is comparable to, and in some instances larger than, those reported for other organizational interventions. Specifically, Guzzo, Jette, and Katzell (1985) reported a mean effect size of 0.44 for all psychologically based interventions, 0.35 for appraisal and feedback, 0.12 for management by objectives, and 0.75 for goal setting on productivity. Kluger and DeNisi (1996) reported a mean effect size of 0.41 for the relationship between feedback and performance. Finally, Neuman, Edwards, and Raju (1989) reported a mean effect size of 0.33 between organizational development interventions and attitudes.

The within-study analyses for training evaluation criteria obtained additional noteworthy results. Specifically, they indicated that comparisons of learning criteria with subsequent criteria (i.e., behavioral and results) showed a substantial decrease in effect sizes from learning to these criteria. This effect may be due to the fact that the manifestation of training learning outcomes in subsequent job behaviors (behavioral criteria) and organizational indicators (results criteria) may be a function of the favorability of the posttraining environment for the performance of the learned skills. Environmental favorability is the extent to which the transfer or work environment is supportive of the application of new skills and behaviors learned or acquired in training (Noe, 1986; Peters & O'Connor, 1980). Trained and learned skills will not be demonstrated as job-related behaviors or performance if incumbents do not have the opportunity to perform them (Arthur et al., 1998; Ford et al., 1992). Thus, for studies using behavioral or results criteria, the social context and the favorability of the posttraining environment play an important role in facilitating the transfer of trained skills to the job and may attenuate the effectiveness of training (Colquitt et al., 2000; Fecteau, Dobbins, Russell, Ladd, & Kudisch, 1992; Tannenbaum, Mathieu, Salas, & Cannon-Bowers, 1991; Tracey et al., 1995; Williams, Thayer, & Pond, 1991).

In terms of needs assessment, although anecdotal information suggests that it is prudent to conduct a needs assessment as the first step in the design and development of training (Ostroff & Ford, 1989; Sleezer, 1993), only 6% of the studies in our data set reported any needs assessment activities prior to training implementation. Of course, it is conceivable and even likely that a much larger percentage conducted a needs assessment but failed to report it in the published work because it may not have been a variable of interest. Contrary to what we expected—that implementation of more comprehensive needs assessments (i.e., the presence of multiple aspects [i.e., organization, task, and person analysis] of the process) would result in more effective training—there was no clear pattern of results for the needs assessment analyses. However, these analyses were based on a small number of data points.

Concerning the choice of training methods for specified skills and tasks, our results suggest that the effectiveness of organizational training appears to vary as a function of the specified training delivery method, the skill or task being trained, and the criterion used to operationalize effectiveness. We highlight the effectiveness of lectures as an example because despite their widespread use (Van Buren & Erskine, 2002), they have a poor public image as a boring and ineffective training delivery method (Carroll, Paine, & Ivancevich, 1972). In contrast, a noteworthy finding in our meta-analysis was the robust effect obtained for lectures, which contrary to their poor public image, appeared to be quite effective in training several types of skills and tasks.

Because our results do not provide information on exactly why a particular method is more effective than others for specified skills or tasks, future research should attempt to identify what instructional attributes of a method impact the effectiveness of that method for different training content. In addition, studies examining the differential effectiveness of various training methods for the same content and a single training method across a variety of skills and tasks are warranted. Along these lines, future research might consider the effectiveness and efficacy of high-technology training methods such as Web-based training.

### *Limitations and Additional Suggestions for Future Research*

First, we limited our meta-analysis to features over which practitioners and researchers have a reasonable amount of control. There are obviously several other factors that could also play a role in the observed effectiveness of organizational training. For instance, because they are rarely manipulated, researched, or reported in the extant literature, two additional steps commonly listed in the training development and evaluation sequence, namely (a) developing the training objectives and (b) designing the evaluation and the actual presentation of the training content, including the skill of the trainer, were excluded. Other factors that we did not investigate include contextual factors such as participation in training-related decisions, framing of training, and organizational climate (Quiñones, 1995, 1997). Additional variables that could influence the observed effectiveness of organizational training include trainer effects (e.g., the skill of the trainer), quality of the training content, and trainee effects such as motivation (Colquitt et al., 2000), cognitive ability (e.g., Ree & Earles, 1991; Warr & Bunce, 1995), self-efficacy (e.g., Christoph, Schoenfeld, & Tansky, 1998; Martocchio & Judge, 1997; Mathieu, Martineau, &

Tannenbaum, 1993), and goal orientation (e.g., Fisher & Ford, 1998). Although we considered them to be beyond the scope of the present meta-analysis, these factors need to be incorporated into future comprehensive models and investigations of the effectiveness of organizational training.

Second, this study focused on fairly broad training design and evaluation features. Although a number of levels within these features were identified a priori and examined, given the number of viable moderators that can be identified (e.g., trainer effects, contextual factors), it is reasonable to posit that there might be additional moderators operating here that would be worthy of future investigation.

Third, our data were limited to individual training interventions and did not include any team training studies. Thus, for instance, our training methods did not include any of the burgeoning team training methods and strategies such as cross-training (Blickensderfer, Cannon-Bowers, & Salas, 1998), team coordination training (Prince & Salas, 1993), and distributed training (Dwyer, Oser, Salas, & Fowlkes, 1999). Although these methods may use training methods similar to those included in the present study, it is also likely that their application and use in team contexts may have qualitatively different effects and could result in different outcomes worthy of investigation in a future meta-analysis.

Finally, although it is generally done more for convenience and ease of explanation than for scientific precision, a commonly used training method typology in the extant literature is the classification of training methods into on-site and off-site methods. It is striking that out of 397 data points, only 1 was based on the sole implementation of an on-site training method. A few data points used an on-site method in combination with an off-site method ( $k = 12$  [3%]), but the remainder used off-site methods only. Wexley and Latham (1991) noted this lack of formal evaluation for on-site training methods and called for a “science-based guide” to help practitioners make informed choices about the most appropriate on-site methods. However, our data indicate that after more than 10 years since Wexley and Latham’s observation, there is still an extreme paucity of formal evaluation of on-site training methods in the extant literature. This may be due to the informal nature of some on-site methods such as on-the-job training, which makes it less likely that there will be a structured formal evaluation that is subsequently written up for publication. However, because of on-site training methods’ ability to minimize costs, facilitate and enhance training transfer, as well as their frequent use by organizations, we reiterate Wexley and Latham’s call for research on the effectiveness of these methods.

### *Conclusion*

In conclusion, we identified specified training design and evaluation features and then used meta-analytic procedures to empirically assess their relationships to the effectiveness of training in organizations. Our results suggest that the training method used, the skill or task characteristic trained, and the choice of training evaluation criteria are related to the observed effectiveness of training programs. We hope that both researchers and practitioners will find the information presented here to be of some value in making informed choices and decisions in the design, implementation, and evaluation of organizational training programs.

## References

- Alliger, G. M., & Janak, E. A. (1989). Kirkpatrick's levels of training criteria: Thirty years later. *Personnel Psychology, 41*, 331-342.
- Alliger, G. M., Tannenbaum, S. I., Bennett, W., Jr., Traver, H., & Shotland, A. (1997). A meta-analysis of relations among training criteria. *Personnel Psychology, 50*, 341-358.
- Arthur, W., Jr., Bennett, W., Jr., & Huffcutt, A. I. (2001). *Conducting meta-analysis using SAS*. Mahwah, NJ: Erlbaum.
- Arthur, W., Jr., Bennett, W., Jr., Stanush, P. L., & McNelly, T. L. (1998). Factors that influence skill decay and retention: A quantitative review and analysis. *Human Performance, 11*, 57-101.
- Arthur, W., Jr., Day, E. A., McNelly, T. L., & Edens, P. S. (in press). Distinguishing between methods and constructs: The criterion-related validity of assessment center dimensions. *Personnel Psychology*.
- Arthur, W., Jr., Tubre, T. C., Paul, D. S., & Edens, P. S. (in press). Teaching effectiveness: The relationship between reaction and learning criteria. *Educational Psychology, 23*, 275-285.
- Blickensderfer, E., Cannon-Bowers, J. A., & Salas, E. (1998). Cross training and team performance. In J. A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 299-311). Washington, DC: American Psychological Association.
- Burke, M. J., & Day, R. R. (1986). A cumulative study of the effectiveness of managerial training. *Journal of Applied Psychology, 71*, 232-245.
- Campbell, J. P. (1971). Personnel training and development. *Annual Review of Psychology, 22*, 565-602.
- Carroll, S. J., Paine, F. T., & Ivancevich, J. J. (1972). The relative effectiveness of training methods—expert opinion and research. *Personnel Psychology, 25*, 495-510.
- Cascio, W. F. (1991). *Costing human resources: The financial impact of behavior in organizations* (3rd ed.). Boston: PWS-Kent Publishing Co.
- Cascio, W. F. (1998). *Applied psychology in personnel management* (5th ed.). Upper Saddle River, NJ: Prentice Hall.
- Christoph, R. T., Schoenfeld, G. A., Jr., & Tansky, J. W. (1998). Overcoming barriers to training utilizing technology: The influence of self-efficacy factors on multimedia-based training receptiveness. *Human Resource Development Quarterly, 9*, 25-38.
- Cohen, J. (1992). A power primer. *American Psychologist, 112*, 155-159.
- Colquitt, J. A., LePine, J. A., & Noe, R. A. (2000). Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology, 85*, 678-707.
- Day, E. A., Arthur, W., Jr., & Gettman, D. (2001). Knowledge structures and the acquisition of a complex skill. *Journal of Applied Psychology, 86*, 1022-1033.
- Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1*, 1-8.
- Dwyer, D. J., Oser, R. L., Salas, E., & Fowlkes, J. E. (1999). Performance measurement in distributed environments: Initial results and implications for training. *Military Psychology, 11*, 189-215.
- Facteau, J. D., Dobbins, G. H., Russell, J. E. A., Ladd, R. T., & Kudisch, J. D. (1992). *Noe's model of training effectiveness: A structural equations analysis*. Paper presented at the Seventh Annual Conference of the Society for Industrial and Organizational Psychology, Montreal, Quebec, Canada.
- Facteau, J. D., Dobbins, G. H., Russell, J. E. A., Ladd, R. T., & Kudisch, J. D. (1995). The influence of general perceptions of the training environment on pretraining motivation and perceived training transfer. *Journal of Management, 21*, 1-25.
- Farina, A. J., Jr., & Wheaton, G. R. (1973). Development of a taxonomy of human performance: The task-characteristics approach to performance prediction. *JSAS Catalog of Selected Documents in Psychology, 3*, 26-27 (Manuscript No. 323).
- Fisher, S. L., & Ford, J. K. (1998). Differential effects of learner effort and goal orientation on two learning outcomes. *Personnel Psychology, 51*, 397-420.
- Fleishman, E. A., & Quaintance, M. K. (1984). *Taxonomies of human performance: The description of human tasks*. Orlando, FL: Academic Press.
- Ford, J. K., Quiñones, M., Sego, D. J., & Speer Sorra, J. S. (1992). Factors affecting the opportunity to perform trained tasks on the job. *Personnel Psychology, 45*, 511-527.
- Gagne, R. M., Briggs, L. J., & Wagner, W. W. (1992). *Principles of instructional design*. New York: Harcourt Brace Jovanovich.
- Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social science research*. Beverly Hills, CA: Sage.
- Goldstein, I. L. (1980). Training in work organizations. *Annual Review of Psychology, 31*, 229-272.
- Goldstein, I. L., & Ford, J. K. (2002). *Training in organizations: Needs assessment, development, and evaluation* (4th ed.). Belmont, CA: Wadsworth.
- Guzzo, R. A., Jette, R. D., & Katzell, R. A. (1985). The effects of psychologically based intervention programs on worker productivity: A meta-analysis. *Personnel Psychology, 38*, 275-291.
- Huffcutt, A. I., & Arthur, W., Jr. (1995). Development of a new outlier statistic for meta-analytic data. *Journal of Applied Psychology, 80*, 327-334.
- Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.
- Industry report 2000. (2000). *Training, 37*(10), 45-48.
- Kaplan, R. M., & Pascoe, G. C. (1977). Humorous lectures and humorous examples: Some effects upon comprehension and retention. *Journal of Educational Psychology, 69*, 61-65.
- Kirkpatrick, D. L. (1959). Techniques for evaluating training programs. *Journal of the American Society of Training and Development, 13*, 3-9.
- Kirkpatrick, D. L. (1976). Evaluation of training. In R. L. Craig (Ed.), *Training and development handbook: A guide to human resource development* (2nd ed., pp. 301-319). New York: McGraw-Hill.
- Kirkpatrick, D. L. (1996). Invited reaction: Reaction to Holton article. *Human Resource Development Quarterly, 7*, 23-25.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*, 254-284.
- Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology, 78*, 311-328.
- Latham, G. P. (1988). Human resource training and development. *Annual Review of Psychology, 39*, 545-582.
- Martocchio, J. J., & Judge, T. A. (1997). Relationship between conscientiousness and learning in employee training: Mediating influences of self-deception and self-efficacy. *Journal of Applied Psychology, 82*, 764-773.
- Mathieu, J. E., Martineau, J. W., & Tannenbaum, S. I. (1993). Individual and situational influences on the development of self-efficacy: Implication for training effectiveness. *Personnel Psychology, 46*, 125-147.
- McGehee, W., & Thayer, P. W. (1961). *Training in business and industry*. New York: Wiley.
- Neuman, G. A., Edwards, J. E., & Raju, N. S. (1989). Organizational development interventions: A meta-analysis of their effects on satisfaction and other attitudes. *Personnel Psychology, 42*, 461-489.
- Noe, R. A. (1986). Trainee's attributes and attitudes: Neglected influences on training effectiveness. *Academy of Management Review, 11*, 736-749.
- Noe, R. A., & Schmitt, N. M. (1986). The influence of trainee attitudes on training effectiveness: Test of a model. *Personnel Psychology, 39*, 497-523.
- Ostroff, C., & Ford, K., J. (1989). Critical levels of analysis. In I. L.

- Goldstein (Ed.), *Training and development in organizations* (pp. 25–62). San Francisco, CA: Jossey-Bass.
- Peters, L. H., & O'Connor, E. J. (1980). Situational constraints and work outcomes: The influence of a frequently overlooked construct. *Academy of Management Review*, 5, 391–397.
- Prince, C., & Salas, E. (1993). Training and research for teamwork in military aircrew. In E. L. Wiener, B. G. Kanki, & R. L. Helmreich (Eds.), *Cockpit resource management* (pp. 337–366). San Diego, CA: Academic Press.
- Quiñones, M. A. (1995). Pretraining context effects: Training assignment as feedback. *Journal of Applied Psychology*, 80, 226–238.
- Quiñones, M. A. (1997). Contextual influences on training effectiveness. In M. A. Quiñones & A. Ehrenstein (Eds.), *Training for a rapidly changing workplace: Applications of psychological research* (pp. 177–199). Washington, DC: American Psychological Association.
- Quiñones, M. A., Ford, J. K., Sego, D. J., & Smith, E. M. (1995). The effects of individual and transfer environment characteristics on the opportunity to perform trained tasks. *Training Research Journal*, 1, 29–48.
- Rasmussen, J. (1986). *Information processing and human-machine interaction: An approach to cognitive engineering*. New York: Elsevier.
- Ree, M. J., & Earles, J. A. (1991). Predicting training success: Not much more than g. *Personnel Psychology*, 44, 321–332.
- Salas, E., & Cannon-Bowers, J. A. (2001). The science of training: A decade of progress. *Annual Review of Psychology*, 52, 471–499.
- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 84, 1–66.
- Severin, D. (1952). The predictability of various kinds of criteria. *Personnel Psychology*, 5, 93–104.
- Sleezer, C. M. (1993). Training needs assessment at work: A dynamic process. *Human Resource Development Quarterly*, 4, 247–264.
- Tannenbaum, S. I., Mathieu, J. E., Salas, E., & Cannon-Bowers, J. A. (1991). Meeting trainees' expectations: The influence of training fulfillment on the development of commitment, self-efficacy, and motivation. *Journal of Applied Psychology*, 76, 759–769.
- Tannenbaum, S. I., & Yukl, G. (1992). Training and development in work organizations. *Annual Review of Psychology*, 43, 399–441.
- 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, 239–252.
- Van Buren, M. E., & Erskine, W. (2002). *The 2002 ASTD state of the industry report*. Alexandria, VA: American Society of Training and Development.
- Warr, P., & Bunce, D. (1995). Trainee characteristics and the outcomes on open learning. *Personnel Psychology*, 48, 347–375.
- Wexley, K. N. (1984). Personnel training. *Annual Review of Psychology*, 35, 519–551.
- Wexley, K. N., & Latham, G. P. (1991). *Developing and training human resources in organizations* (2nd ed.). New York: Harper Collins.
- Wexley, K. N., & Latham, G. P. (2002). *Developing and training human resources in organizations* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.
- Whitener, E. M. (1990). Confusion of confidence intervals and credibility intervals in meta-analysis. *Journal of Applied Psychology*, 75, 315–321.
- Williams, T. C., Thayer, P. W., & Pond, S. B. (1991, April). *Test of a model of motivational influences on reactions to training and learning*. Paper presented at the Sixth Annual Conference of the Society for Industrial and Organizational Psychology, St. Louis, Missouri.
- Wolf, F. M. (1986). *Meta-analysis: Quantitative methods for research synthesis*. Newbury Park, CA: Sage.
- Zemke, R. E. (1994). Training needs assessment: The broadening focus of a simple construct. In A. Howard (Ed.), *Diagnosis for organizational change: Methods and models* (pp. 139–151). New York: Guilford Press.

Received March 7, 2001

Revision received March 28, 2002

Accepted April 29, 2002 ■

### Wanted: Your Old Issues!

As APA continues its efforts to digitize journal issues for the PsycARTICLES database, we are finding that older issues are increasingly unavailable in our inventory. We are turning to our long-time subscribers for assistance. If you would like to donate any back issues toward this effort (preceding 1982), please get in touch with us at [journals@apa.org](mailto:journals@apa.org) and specify the journal titles, volumes, and issue numbers that you would like us to take off your hands.

Copyright of Journal of Applied Psychology is the property of American Psychological Association and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.



Copyright of Journal of Applied Psychology is the property of American Psychological Association and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.