Journal of Applied Psychology
2005, Vol. 90, No. 4, 692–709

0021-9010/05/$12.00 DOI: 10.1037/0021-9010.90.4.692

# A Meta-Analytic Review of Behavior Modeling Training

Paul J. Taylor
Chinese University of Hong Kong and University of Waikato

Darlene F. Russ-Eft
Oregon State University

Daniel W. L. Chan
Chinese University of Hong Kong

A meta-analysis of 117 studies evaluated the effects of behavior modeling training (BMT) on 6 training outcomes, across characteristics of training design. BMT effects were largest for learning outcomes, smaller for job behavior, and smaller still for results outcomes. Although BMT effects on declarative knowledge decayed over time, training effects on skills and job behavior remained stable or even increased. Skill development was greatest when learning points were used and presented as rule codes and when training time was longest. Transfer was greatest when mixed (negative and positive) models were presented, when practice included trainee-generated scenarios, when trainees were instructed to set goals, when trainees' superiors were also trained, and when rewards and sanctions were instituted in trainees' work environments.

*Keywords:* training, behavior modeling training, training methods, managerial training, supervisory training

Behavior modeling training (BMT) has become one of the most widely used, well-researched, and highly regarded psychologically based training interventions. The approach, based on Bandura's (1977) social learning theory, differs from other training methods with its emphasis on (a) describing to trainees a set of well-defined behaviors (skills) to be learned, (b) providing a model or models displaying the effective use of those behaviors, (c) providing opportunities for trainees to practice using those behaviors, (d) providing feedback and social reinforcement to trainees following practice, and (e) taking steps to maximize the transfer of those behaviors to the job (Decker & Nathan, 1985; Goldstein & Sorcher, 1974; Robinson, 1982). Although other training approaches often include one or more of these components, BMT emphasizes the importance of including them all.

BMT has been applied in the development of supervisory, communications, sales, and customer service skills training programs by both large organizations and leading international training companies. One of the largest training firms using BMT for teaching supervisory skills has estimated that over 6 million managers in over 8,000 organizations have taken its courses (Wexley & Latham, 2002). Recently, BMT has been extended to a broader range of applications, including cross-cultural skills (e.g., Harrison, 1987, 1992) and technical skills (e.g., Compeau & Higgins, 1995; Gist, Rosen, & Schwoerer, 1988; Gist, Schwoerer, & Rosen, 1989; Simon & Werner, 1996).

Here we provide a meta-analytic review of BMT research, considering its impact on a comprehensive set of training outcomes, and under a variety of conditions related to training design, such as how learning points and models are presented, features of behavioral rehearsal, and steps taken to enhance transfer. Next we describe the theory and practice of BMT, followed by a brief review of previous BMT research and an explanation of our aims for the present research.

## The Theory and Practice of BMT

Social learning theory (Bandura, 1977), which is the foundation of BMT, emphasizes four component processes: attentional, retentional, reproduction, and motivational processes. Within the context of BMT, attentional processes concern trainees observing modeling stimuli (e.g., a videotape of someone depicting the skills or desired behavior). The extent to which a learner attends to the modeled behavior is thought to be influenced by (a) characteristics of how the modeled behaviors are displayed, for example, sequencing of behaviors from least to most difficult, labeling of key behaviors; (b) characteristics of the key behaviors being modeled, for example, how distinctively they are presented; (c) characteristics of the model, for example, learners' perceptions of the model's expertise, and similarities between learners and the model such as

---

Paul J. Taylor, Department of Psychology, Chinese University of Hong Kong, Hong Kong, and Department of Psychology, University of Waikato, Hamilton, New Zealand; Darlene F. Russ-Eft, Department of Adult Education and Higher Education Leadership, Oregon State University; Daniel W. L. Chan, Department of Psychology, Chinese University of Hong Kong.

This project was supported in part by grants from the Faculty of Social Science at the Chinese University of Hong Kong and the Department of Psychology at the University of Waikato. We thank Michael Burke and Frank Schmidt for helpful comments on an earlier version of this article; Kristi Andes, Jane Ho, Vivian Lun, Iris Ng, and Rosanna Yip for assistance in obtaining and coding studies; and those who identified and contributed unpublished studies for this article, particularly Beth Blickensderfer, William Byham, Phillip Decker, James Robinson, Eduardo Salas, and Cynthia Woods.

Correspondence concerning this article should be addressed to Paul J. Taylor, Department of Psychology, University of Waikato, Private Bag 3105, Hamilton, New Zealand. E-mail: ptaylor@waikato.ac.nz

in sex, race, and age; and (d) characteristics of the learner, for example, capabilities and arousal level (Decker & Nathan, 1985).

Whereas attentional processes are believed to underlie the effective transfer of observed stimuli (e.g., modeled behaviors) to short-term memory, retentional processes are considered necessary for learning to transfer to long-term memory. Retentional processes primarily concern *symbolic coding*, in which learners organize behaviors presented during modeling into symbols that facilitate storage and retrieval of those behaviors (Decker & Nathan, 1985). In BMT, retentional processes can be facilitated by using learning points that aid symbolic coding and by encouraging trainees to engage in *symbolic rehearsal* (i.e., mental practice) of how they plan to use modeled behaviors in practice.

Reproduction and motivational processes occur in BMT as trainees practice the skills previously presented through modeling and apply them to the posttraining environment. Practice during BMT, referred to as *behavioral rehearsal* or *skill practice*, includes feedback from other trainees and/or the trainer, which serves not only a correctional function, when trainees fail to use key behaviors successfully, but also a motivational function, through an explicit emphasis on social reinforcement (i.e., praise) when trainees effectively use the newly learned skills (Goldstein & Sorcher, 1974). Motivational processes are facilitated further in BMT with its emphasis on transfer of training to the posttraining environment, primarily through establishing reinforcements for the use of newly learned skills.

The typical sequence of activities with BMT programs is the provision of a description of skills–behaviors to be learned, prior to, or along with, modeling, and then practice with feedback. In the case of teaching complex skills required over a variety of related job tasks, such as interpersonal supervisory skills, training is usually organized into 4 to 12 modules, each between 2 and 4 hr in length, with each module focusing on a particular interpersonal task (e.g., discussing a performance problem with an employee), and including the full sequence of BMT components described above.

A similar sequence of activities is typically used when BMT has been applied to teaching other skills, although often with fewer modules. For example, based on a literature review on cross-cultural differences between the United States and Japan, Harrison (1992) developed a BMT program to teach cross-cultural skills for managers from the United States working in military installations in Japan. The training included two modules covering two dimensions of cultural differences that emerged from a review of the literature. Each module included the presentation of learning points, a videotaped model depicting those learning points, behavioral rehearsal, and feedback and social reinforcement. In teaching trainees how to use a financial software package, Gist and her colleagues (Gist et al., 1989) used BMT by presenting and demonstrating to trainees, through a video display, the steps required to complete various tasks within the software program. As the video was paused between steps, trainees practiced each step (behavioral rehearsal) and received computer feedback on their performance. Similarly, Howard (1992) implemented BMT to teach the operation of a photocopying machine; trainees were presented with learning points and live or video modeling of those learning points, followed by a practice session (behavioral rehearsal) and social reinforcement.

## Prior Research and Aims of the Present Study

Results of BMT research have been summarized through five narrative reviews (Decker & Nathan, 1985; Mayer & Russell, 1987; McGhee & Tullar, 1978; Robinson, 1982; Russ-Eft & Zenger, 1995) and as part of two quantitative reviews of managerial training methods, in which BMT was one of a variety of methods considered (Burke & Day, 1986; Falcone, 1985). Together, these previous reviews have provided a sound basis for understanding BMT processes and examples of the breadth of contexts in which BMT has been successfully applied. The two quantitative reviews, however, had only five BMT studies available for inclusion at that time. Thus, the primary aim of the present investigation was to provide a substantially more comprehensive, quantitative review of BMT effects on training outcomes than has previously been published, pooling larger sets of both published and unpublished studies. Accurate estimates of training effects are important for both researchers and practitioners in that they provide a basis for comparing the effects of alternative interventions, for conducting power analyses in future research, and for making training utility estimates.

Previous reviews have relied heavily on results of BMT studies reported in a 1976 issue of *Personnel Psychology* (Burnaska, 1976; Byham, Adams, & Kiggins, 1976; Goldstein & Sorcher, 1974; Latham & Saari, 1979; Moses & Ritchie, 1976; Smith, 1976), with uniformly positive and large training effects reported for both behavioral changes and results of changed job behavior (e.g., reduced absenteeism resulting from improved supervisory skills). However, more recently published studies of BMT have not been uniformly positive. For example, some recent studies have failed to find significant changes in job behavior, despite evidence of learning (e.g., May & Kahnweiler, 2000; Russell, Wexley, & Hunter, 1984; Werner, O'Leary-Kelly, Baldwin, & Wexley, 1994). Thus, we expected that a more comprehensive, quantitative review of BMT was likely to result in more conservative estimates of BMT effects on changes in job behavior and results of changed behavior.

A related aim was to determine the effectiveness of BMT in teaching different types of skills. Whereas most of the early studies of BMT involved interpersonal skills, such as supervisory and sales skills, more recent applications have included the teaching of technical skills, such as the use of computer software programs. Differences in skills taught using BMT might influence training effect sizes on learning measures because these are influenced by baseline (pretraining) levels of performance. Initial levels of performance are likely to be higher, and thus effect sizes smaller, when training is intended to enhance existing task performance (e.g., interpersonal skills training) than when trainees are taught to perform a new task, such as how to operate a computer system or software package with which they are unfamiliar.

In addition to establishing BMT effects on various training outcomes, we also investigated the stability of BMT effects over time. Little is known about the extent to which BMT outcomes are maintained over time. A previous meta-analytic review of research on the retention of learning and on-the-job behavior change following training (Arthur, Bennett, Stanush, & McNelly, 1998) found evidence of substantial decay in training effects over time with lack of use, although that meta-analysis included no studies using the behavior modeling method. We anticipated less decay in skills–behavior over time in the case of BMT, however, as one of the explicit goals of BMT is to maintain training effects on job behavior over time (Decker & Nathan, 1985; Goldstein & Sorcher, 1974). Furthermore, the context of most BMT applications has been teaching skills that are clearly related to trainees' current or immediate jobs, and so there should be ample opportunity for trainees to continue using skills learned in training.

## Characteristics of BMT Design

Our second major aim was to determine the training design characteristics associated with greater BMT effects. Researchers have suggested, and in some cases tested, variations in how learning points and models can be presented to trainees to facilitate attentional and retentional processes, how reproduction processes can be enhanced as trainees prepare for behavioral rehearsals, and how motivational processes can be enhanced to maximize BMT transfer. Using meta-analytic methods, we explored the efficacy of such training design features by (a) comparing effect size estimates where there is between-studies variation in how BMT has been implemented, and (b) calculating effect size estimates for particular design features from primary studies in which these design features have been experimentally manipulated, that is, a within-studies analytic approach.

Next we describe the features of BMT design that we have focused on in this meta-analysis. We derived these variables from a review of the BMT literature, although we later eliminated some variables of interest because too few studies were available for meaningful meta-analyses. We focus here on only those variables that we were able to analyze.

### *Learning Points*

An essential component of BMT is that desired behaviors (skills) are described to trainees, so as to facilitate trainees' symbolic coding, and this is typically achieved through providing a list of learning points (also referred to as *codes*) to trainees, prior to or during modeling. Goldstein and Sorcher (1974) suggested that BMT is more effective than alternative training techniques, in part, because desired behaviors are more specifically defined for trainees. Similarly, Decker and Nathan (1985) claimed that learning points facilitate both attentional and retentional processes, and so we expected that the inclusion of specific learning points within BMT would result in greater learning.

Researchers have also considered the most effective form in which to present learning points to trainees. Initiated first by Decker (1980), BMT researchers have compared the presentation of learning points to trainees as rules to be followed ("rule codes"), for example, "Listen and respond with empathy to reduce defensiveness," versus descriptions or summaries of the model's behavior, for example, "listened empathetically." Rule codes have been found to be superior to learning points presented as descriptions or summaries of behavior in leading to generalization of skills (Decker & Nathan, 1985), as measured through trainees performance in simulated role-play situations, leading us to expect that effect sizes for learning outcomes would be greater when learning points are presented as rules rather than descriptions or summaries of behavior. We also explored the effects of two additional variables to do with learning points on learning outcomes: (a) whether studies explicitly mentioned the use of retention aids (Decker & Nathan, 1985), such as cards summarizing learning points, and (b) whether learning points were presented along with the modeling display, a strategy advocated as a means of aiding attentional processes by making modeled behaviors more distinctive (Latham & Saari, 1979; Mann & Decker, 1984).

### *Models*

Perhaps the most interesting and researched variable concerning the modeling display component of BMT is the use of positive-only versus mixed (negative and positive) models. Baldwin (1992) argued that displaying mixed models (i.e., both negative and positive models) can more effectively help trainees generalize learning to novel situations (i.e., apply skills in a scenario different from that which was modeled) than displaying positive-only models for two reasons: (a) Generalization is positively associated with greater stimulus variability, which is provided by mixed models, and (b) trainees can better "unlearn" undesirable behaviors when these behaviors, along with desirable behaviors, are depicted. Although Baldwin's study showed mixed models to be superior to positive-only models when trainees were given the opportunity to generalize newly learned skills to a novel task within his experiment, little is known about the comparative efficacy of mixed versus positive-only models in actual work-related training settings and with respect to other training outcomes, particularly training transfer.

### *Behavioral Rehearsal*

Prior to the behavioral rehearsal component of BMT, trainees can be encouraged to mentally rehearse how they will use newly learned skills (i.e., symbolic rehearsal) or they may be coached by another trainee or the trainer in order to explicitly plan how skills will be used. These additions to BMT have been hypothesized to improve retention and performance in the behavioral rehearsal component of training (Decker & Nathan, 1985), and in the case of symbolic rehearsal, evidence has supported this claim (Decker, 1982). Prior meta-analytic evidence, outside of the field of BMT, also suggests that mental rehearsal can lead to greater task performance, particularly for cognitive tasks (Driskell, Copper, & Moran, 1994). Thus, we assessed whether skill development is enhanced with symbolic rehearsal and coaching prior to the behavioral rehearsal.

We also compared training effects on job behavior of studies that had trainees practice their own work-related scenarios with studies using only trainer-provided scenarios. Having trainees develop at least some of the scenarios that they practice in training has been advocated as a means of enhancing training transfer (Robinson, 1982; Wexley & Latham, 2002) and is consistent with the training transfer principle of *identical elements*, that is, enhancing the similarity between the training and work environments. Thus, we expected larger BMT effects on job behavior when the behavioral rehearsals included trainees practicing at least some scenarios that they developed themselves.

### *Hours of Training*

Substantial practice, to a point of overlearning, has been considered an important aspect of BMT (Decker & Nathan, 1985), and some have even argued that, to achieve transfer of training to the work environment, more practice is required in BMT than has typically been acknowledged (Campbell & Kuncel, 2001). Support for the relationship between practice time in training and retention of newly learned skills can also be found outside of the BMT literature. In a meta-analysis of studies involving physical and cognitive skill development (Driskell, Willis, & Copper, 1992), overlearning was associated with greater retention of newly learned cognitive skills and, to a lesser extent, physical skills.

Although too few BMT studies have included details of how much training time was dedicated specifically to practice, many studies indicated the total number of hours of training time, which we used as a proxy for amount of practice trainees received for the purpose of determining relationships between practice time and training effect sizes. We assumed that the number of hours of training and number of hours of practice are likely to have been at least moderately correlated, given that a relatively large proportion of training time within BMT is dedicated to practice, and longer programs often involve multiple, related modules, each including a behavioral rehearsal component.

### Transfer Enhancers

Finally, with BMT's emphasis on transfer of training to the workplace, researchers have proposed a variety of design features to maximize transfer. We focus here on three such strategies that have been suggested: (a) having trainees set goals in training concerning how they intend to apply newly learned skills on the job (Latham & Saari, 1979; Russell et al., 1984); (b) training the superiors of trainees (Latham & Saari, 1979; Parry & Reich, 1984); and (c) introducing rewards and sanctions in the workplace for, respectively, trainees' using or failing to use newly learned skills on the job (Decker & Nathan, 1985; Goldstein & Sorcher, 1974; Latham & Saari, 1979; Russell et al., 1984).

## Method

Meta-analysis was used in order to establish effect size estimates for relevant training outcomes and to compare training effects across features of both training and evaluation design.

### Study Selection and Compilation

Study selection criteria were as follows.

1. The evaluation had to be of a BMT intervention used with an adult (age 18 or older) population. Authors must have identified the training as *behavior modeling* or *behavior role modeling*, and studies had to include key behavior modeling components of (a) describing and (b) modeling skills (key behaviors) for trainees followed by (c) practice with (d) feedback.[1]

2. The study must have either (a) indicated an effect of BMT, alone, on learning, job behavior, or results of trainees' job behavior, in comparison to a no-training assessment (either control group, pretraining measure, or both), or (b) presented a comparison of two or more variations on the implementation of BMT (e.g., learning points presented vs. not presented to trainees). Evaluations using only reaction questionnaires were excluded in the present investigation because they fail to fit the traditional concept of a "training effect," in that posttraining reactions to training cannot be meaningfully compared with reactions of those who have not received training. Studies adding behavior modeling to another training method were excluded if the full, independent effects of BMT could not be disentangled from the effects of the other training method. Studies utilizing only trainees' own self-ratings were also excluded in light of past research suggesting that self-ratings of performance are typically inflated (Arnold & Davey, 1992; Carless & Roberts-Thompson, 2001; Church, 1997; Furnham & Stringfield, 1994; Harris & Schaubroeck, 1988; Meyer, 1980; Thornton, 1980). Studies relying exclusively on "retrospective" pretests (i.e., neither true pretests nor control groups) were also excluded.

3. Sufficient information had to be presented in order to calculate effect sizes.

A comprehensive search was conducted to find both published and unpublished studies that met these inclusion criteria. Relevant computer databases (PsycINFO, ABI/INFORM, Educational Resources Information Center) were searched using keywords combining *training* with both American and British spellings of *behavior modeling*, *behavioral modeling*, or *behavior role modeling*, as were leading research journals in the training, industrial–organizational psychology, and management fields. Studies were also identified from books and chapters on BMT (Decker & Nathan, 1985; Robinson, 1982; Russ-Eft, 1997; Russ-Eft & Zenger, 1995) and from previously published reviews and meta-analyses of organizational training. Further studies were identified from references cited in studies found through this search process. In order to obtain unpublished BMT evaluations, known BMT researchers were contacted, as were the American Society for Training and Development and training firms that advertise regularly in *Training and Development Journal* and *Training* magazine.

In total, 171 BMT studies were located, of which 46 were later excluded: 26 because they lacked sufficient information for effect size calculations (e.g., Goldstein & Sorcher, 1974); 6 because they included neither pretests nor a relevant comparison group (e.g., Johnson & Marakas, 2000); 6 because training effects were measured exclusively through trainees' own self-ratings (e.g., Nunns & Bluen, 1992); 4 because, although they were described by their authors as based on behavior modeling, one or more of the four key components of behavior modeling was not included (e.g., Royster, 1981); 3 because they were implemented with children (e.g., Renaud & Stolovitch, 1988); and 1 because a pure, BMT-only condition could not be isolated (Davis & Mount, 1984).

After removing duplicate studies (e.g., doctoral dissertations that were later published), a total of 119 independent studies contributed 279 effect sizes to this investigation: 36 published studies, 60 unpublished training reports, and 23 unpublished doctoral dissertations. Of the 119 studies, 94 were conducted in work organizations. The remaining 25 studies involved nonwork samples (e.g., students) and contributed effect sizes to only learning and attitudinal outcome criteria. The skills taught in most (108 of the 119) studies were in the nature of interpersonal communication: 78 concerning supervisory skills training and 30 concerning interpersonal communication skills for nonsupervisory staff (e.g., team skills, assertiveness, customer service, cross-cultural communication, counseling–active listening). The remaining 11 studies taught technical skills, of which 8 concerned computer-related tasks. Forty of the 119 studies included control groups receiving no training, although some of the other studies were also

---

[1] Three unpublished reports were not explicitly identified as BMT in written reports but were identified verbally by one of their authors as behavior modeling, and our inspection revealed that they contained the essential components of BMT. Studies comparing alternative means of presenting learning points or models to trainees on performance in subsequent behavioral rehearsals were included, even if the behavioral rehearsal did not include feedback. These studies only contributed to analyses of alternative features of BMT design.

of an experimental design.[2] In almost half of the control group studies, participants were randomly assigned to training and no-training control conditions. Sample sizes of each study (including control group members) ranged from 5 to 271, with a mean of 37. Sources of studies included in this analysis are preceded by an asterisk in the reference list.

## *Coding of Studies*

Studies were coded as meeting the study inclusion criteria, and those that were included were then coded as to publication source; setting (organizational vs. nonorganizational, industry, level of trainees); skills taught; aspects of study design (control group, random assignment, use of pretests, time between training and posttraining measures); outcome variables assessed (and source of ratings); and implementation variables, including whether learning points were presented as rules versus descriptions or summaries of behavior, use of retention aids, display of learning points with modeling videos, whether models were positive-only versus mixed (negative and positive), use of symbolic rehearsal and coaching prior to behavioral rehearsal, use of trainee-generated practice scenarios, hours of training, use of goal setting, whether trainees' superiors were trained, and whether rewards/sanctions were established in trainees' work environment for use/nonuse of newly learned skills. Other features of BMT design, such as live versus videotaped models and use of trainees' superiors versus others as trainers, were initially coded as well, but too little variation across studies on these design features (e.g., insufficient numbers of studies using live models and trainees' superiors as trainers) prohibited meaningful comparisons, and so these potential moderator variables were dropped. For each outcome variable measured in a study, the number of participants in training and control groups was recorded, effect sizes were calculated, and reliability coefficients and pretest–posttest correlations were noted.

Three graduate psychology students were trained as cocoders. Training consisted of cocoders and Paul J. Taylor independently coding a set of six studies, followed by discussion of disagreements, including both categorical differences and differences in effect size calculations of greater than .005. After two iterations of this process, all coders achieved independent agreement with Paul J. Taylor for each variable coded in at least five of the six studies in the set (i.e., at least 83% agreement), and training of cocoders was concluded. After training, each remaining study was independently coded by Paul J. Taylor and one of the three trained cocoders. Posttraining percentage agreement between Paul J. Taylor and cocoders exceeded 90% for all variables except three: skills taught (75%), study design (88%), and effect sizes (74%). Final coding decisions on disagreements occurring after training were made by Paul J. Taylor.

## *Categorization of Training Effects*

Training effects were categorized into one of six outcome criteria: first, declarative knowledge (e.g., typically written tests using multiple-choice format); second, procedural knowledge–skills (skills assessed in simulation tasks or through paper-and-pencil situational judgment tests)[3]; third, training-related attitudes (e.g., self-efficacy); fourth, job behavior; and two results of trainee behavior: workgroup productivity[4] (the fifth outcome) and workgroup climate (the sixth outcome)—both outcomes assessed in many evaluations of BMT supervisory training. Other possible measures of learning suggested by Kraiger, Ford, and Salas (1993), such as mental structures, were originally considered but later dropped because too few studies included them for meaningful meta-analysis.

## *Notation of Training Effects*

Relevant training effects from each study were categorized into one of the six training outcome criteria listed above. In some studies, more than one measure was used for a single evaluation criterion, such as two forms of written tests, performance in multiple simulations, or multiple rating sources for behavioral and results measures. In these cases, multiple effect sizes for a particular criterion construct were averaged. Similarly, for studies including multiple variations of BMT, for example, one condition using positive models and another condition using mixed models, and provided that each condition contained the minimal requirements of BMT defined earlier, effect sizes established for each condition were averaged for analyses involving estimates of BMT training on outcomes and were treated separately for analyses of that particular BMT design feature. In some studies, both immediate and delayed posttests were administered, and in these cases, effect sizes were calculated for each, and these were considered separately for an analysis of differences in immediate versus delayed training effects but were averaged for other analyses.

All effect sizes were based on standardized differences in performance of individuals receiving BMT versus a no-training comparison (e.g., control group or pretest), with the exception of within-studies analyses comparing features of BMT design (e.g., two or more variations of BMT). Some studies compared two or more different training methods, including a no-training control group, and in these cases, effect sizes for BMT versus no training were calculated for only the behavior modeling group as compared with the no-training control group. Similarly, for studies that experimentally compared two or more different training methods without a no-training control group but that included pretest measures, effect sizes for BMT versus no training were based on the pretest–posttest comparison of only the behavior modeling group (or groups), that is, treated as a single-group, pretest–posttest design.

Training effect sizes were calculated based on new procedures suggested by Carlson and Schmidt (1999) in order to minimize biases that result from different evaluation designs: (a) the upward bias in studies lacking control groups, due to control group effects (i.e., improvements in performance over time unrelated to the training, such as the effect of pretesting on posttest scores), and (b) the downward bias in control group studies lacking pretest measures, due to treatment-by-subject interactions that result in inflated posttraining standard deviations. Following recommendations by Carlson and Schmidt (1999) and Morris and DeShon (2002), we calculated training effect sizes in the following manner.

When means and standard deviations were reported in primary studies, effect sizes were calculated as follows.

*Posttest only with control group designs.* Effect sizes were calculated as the mean difference between trained and control groups divided by the control group standard deviation.

*Pretest–posttest with control group designs.* Effect sizes were computed in three stages. First, an effect size was computed for the trained group, using the mean difference between posttest and pretest for the trained group, divided by the trained group pretest standard deviation.

---

[2] Studies with pretests and posttests, along with control groups that only received an alternative training approach (e.g., lecture, computer-assisted instruction), were coded as not having control groups in our analyses of BMT effects (vs. no training), because, without a no-training control group, we were only able to use pretest and posttest measures associated with the study's BMT condition to calculate the BMT versus no-training effect.

[3] A small number of studies have included measures of both behavioral reproduction (in which the trainees are evaluated in using trained skills in the same, or a closely related, situation to that demonstrated in the model) and skill generalization (in which the evaluation involves a situation that is different from the situation demonstrated in the model). Too few studies included behavioral reproduction to meaningfully analyze this training outcome, and so it was dropped.

[4] Surprisingly, studies assessing the effect of BMT on productivity rarely used objective measures of workgroup productivity (e.g., sales or production figures), and often where objective measures were reported, statistics necessary for the computation of effect sizes (e.g., standard deviations) were missing.

Next, an effect size was computed for the control group, using the mean difference between posttest and pretest for the control group, divided by the control group pretest standard deviation. Finally, the effect size for the control group was subtracted from the effect size for the trained group.

*Single group pretest–posttest designs.* Effect sizes were calculated as the mean difference between posttest and pretest scores, divided by the pretest standard deviation. In order to remove bias in these effect sizes due to gains in posttest scores attributable to improvements in performance unrelated to the training (e.g., pretesting), control group effects were subtracted from these single-group, pretest–posttest effect sizes (prior to correction for attenuation) using control group effect size estimates for each training outcome variable provided by Carlson and Schmidt (1999, Table 4): knowledge measures, $d_{\text{control}} = 0.191$; behavioral measures, $d_{\text{control}} = 0.121$; and outcome measures, $d_{\text{control}} = 0.101$. No correction for control group effects in attitude measures was made because Carlson and Schmidt found a near-zero control group effect for this dependent variable.

When means and standard deviations were not reported for studies with control groups, but test statistics (or precise–obtained probability values) were, conversion formulas described by Hunter and Schmidt (1990) and Glass, McGaw, and Smith (1981) were used to calculate effect sizes. For within-subject studies, and for which only correlated $t$ values or repeated-measures $F$ values were reported, the conversion formula provided by Dunlap, Cortina, Vaslow, and Burke (1996) was applied in order to reflect within-subject effects in a comparable metric to effect sizes from between-subjects studies. Effect sizes for single-group pretest–posttest studies derived from test statistics were also corrected for control group effects using the procedure for single-group pretest–posttest designs described above.

Studies that did not report means and standard deviations also failed to report pretest–posttest correlations, and so effect size calculations for these studies were based on an estimation of the pretest–posttest correlation, using the average pretest–posttest correlation reported for the same criterion in other BMT studies, as suggested by Dunlap et al. (1996). These sample-weighted mean pretest–posttest correlations used for imputation are reported in Table 1.

## Removal of Outliers

Outlier effect sizes were identified and removed from analyses according to Huffcutt and Arthur's (1995) sample-adjusted meta-analytic deviance (SAMD) statistic. Their procedure was adapted to the present study by calculating sampling error variances according to the appropriate between-subjects or within-subject formula (see the discussion of sampling variances in the *Meta-Analytic Procedure* section below). Scree plots of SAMD values were generated separately for each of the six training outcomes.

Visual inspection of scree plots identified 11 of the 279 effect sizes (3.9%) as outliers. Some outlier effect sizes were of learning measures, where in at least some cases, the training was aimed at teaching entirely new skills to trainees, such as use of computer software that trainees were previously unfamiliar with (vs. training to improve existing skills, such as interpersonal skills), and so very low trainee pretest scores or control group members' scores would have been expected. For example, Simon and Werner (1996) reported (uncorrected) effect sizes for BMT on declarative knowledge and procedural knowledge and skills of 6.3 and 18.1, respectively. In other cases, a misprint of statistics seemed plausible. For example, Orpen (1985) reported a very large effect size (6.3) for BMT on district sales figures, but because he only reported its significance at the .05 level, while reporting other, smaller effects within the same study at more stringent critical probability values (.01 and .001), a misprint (e.g., misplaced decimal point) seems plausible.

The 11 outlier effect sizes were from 8 different studies, and because most studies contributed multiple effect sizes, removal of these 11 outlier effect sizes resulted in the loss of only 2 studies, leaving a total of 117 studies, and 268 effect sizes, for analysis.

## Meta-Analytic Procedure

Population effect size estimates were calculated by computing sample-size weighted mean effect sizes, in which studies were corrected individually for criterion unreliability, using the Hunter and Schmidt (1990) approach. Effect sizes were corrected individually for unreliability using alpha coefficients, as these were the most commonly reported reliabilities for all criteria. Using alpha coefficients to correct effect sizes corrects only one of multiple sources of unreliability present (F. L. Schmidt & Hunter, 1996), and so effect sizes should be considered conservative estimates of population effects. In cases in which studies lacked criterion reliability information, the sample-weighted mean reliability based on other BMT studies for that criterion (see Table 1) was imputed. Effect sizes were also corrected for the bias associated with studies with small samples (Hunter & Schmidt, 1990, pp. 281–282).

Sampling variance estimates were computed for each study individually, because between-subjects and within-subject studies require different sampling variance estimation procedures. Sampling variances (of uncorrected effect sizes) for posttest-only with control group studies were computed using Hunter and Schmidt's (1990, p. 283) formula for between-subjects studies, whereas sampling variances (of uncorrected effect sizes) for both single-group pretest–posttest studies and pretest–posttest with control group studies were calculated using their formula for within-subject studies (Hunter & Schmidt, 1990, p. 383). For pretest–posttest with control group studies, sampling variances were computed separately for training and control group effect sizes (using the within-subject sampling variance formula) and then summed to compute the total sampling variance for each study as recommended by Morris and DeShon (2002). Pretest–posttest correlations, which are required for computation of within-subject sampling variances, were obtained from source studies reporting these statistics (see Table 1).

Sampling error variances of corrected effect sizes were computed according to Hunter and Schmidt (1990, pp. 307–308), in which each study's

Table 1
*Sample-Weighted Mean Pretest–Posttest Correlations and Reliabilities*

| | Pretest–posttest correlation | | | Reliability | | |
|---|---|---|---|---|---|---|
| Evaluation criterion | Mean $r_{\text{pre–post}}$ | $k$ | $N$ | Mean $r_{yy}$ | $k$ | $N$ |
| Declarative knowledge | .67 | 1 | 68 | .76 | 2 | 191 |
| Procedural knowledge–skills | .60 | 1 | 133 | .85 | 7 | 558 |
| Attitudes | .60 | 32 | 749 | .78 | 42 | 1,150 |
| Job behavior | .82 | 21 | 947 | .86 | 47 | 2,509 |
| Results: workgroup productivity | .44 | 22 | 924 | .89 | 35 | 1,515 |
| Results: workgroup climate | .43 | 21 | 940 | .77 | 35 | 1,533 |

*Note.* Reliabilities are sample-weighted mean alpha coefficients. $k$ = number of studies; $N$ = total sample size.

sampling variance was divided by the reliability of the dependent measure. Sampling variances of corrected effect sizes were used to correct variances in corrected effect sizes, which, in turn, were used to calculate corrected standard deviations and lower bound values of 90% credibility intervals around corrected sample-weighted mean effect sizes ($\delta$).

### *Moderator Analyses*

Moderator analyses of training effects were performed using both between- and within-studies strategies. Relationships between effect sizes and categorical moderators were assessed using the traditional between-studies approach of breaking studies down by hypothesized moderators. Similarly, we report within-studies effect size estimates for studies that experimentally compared multiple conditions relevant to BMT, such as two variations of BMT design (e.g., positive-only vs. mixed models). Meta-analysis of such experiments comparing two treatment conditions is similar to traditional analyses comparing treatment with control conditions, except that effect size estimates represent the effect of one treatment compared with the other rather than the effect of receiving versus not receiving a treatment. For both between-studies and within-studies analyses, we established a minimum of three studies necessary for each subset of studies.

Relationships between effect sizes and continuously measured moderators, such as the length of time between training and posttests and the number of hours of training, were assessed using weighted least squares (WLS) regression, in which (uncorrected) effect sizes were regressed on each continuously measured moderator, and studies were weighted by their sample sizes. We set a minimum of 10 studies as necessary for performing WLS.

## Results

BMT effects on training outcomes, broken down by methodological features of studies (use of control groups and random assignment to groups) are presented in Table 2. Results outcomes are presented separately for workgroup productivity and workgroup climate—both relevant for only BMT supervisory training.

Focusing first on results for learning outcomes (see Table 2), mean population effect size estimates ($\delta$) for BMT on measures of declarative knowledge and procedural knowledge–skills showed the largest training effects, with trainees improving by slightly more than 1 standard deviation. Even after we removed variance due to sampling error, substantial residual variance remained (as indicated by large corrected standard deviations of $\delta$, small percentages of variance accounted for, and low 90% lower bound credibility values). Distributions of effect sizes for both declarative knowledge and procedural knowledge–skills were highly skewed, with virtually all study effect sizes positive and some as large as 4 standard deviations, and so lower credibility values have to be interpreted cautiously. In sum, BMT effects on knowledge and skills measures are generally large and vary considerably from study to study. BMT effects on attitudes are more modest, with an average change of one third of a standard deviation, and they are also more stable across studies. Training effects on job behavior (see Table 2) are substantially smaller than on knowledge and skills, averaging just over one fourth of a standard deviation, with

Table 2
*Behavior Modeling Training Effects on Training Outcomes by Use of Control Groups and Random Assignment*

| Subset of studies | $N$ | No. of studies | Mean $d$ | $\delta$ | Corrected $SD$ of $\delta$ | % variance accounted for | Lower bound 90% credibility value |
|---|---|---|---|---|---|---|---|
| Declarative knowledge | 995 | 14 | 1.05 | 1.20 | 1.34 | 3.2 | −.51 |
| Studies with control groups | 523 | 8 | 0.98 | 1.12 | 1.49 | 3.8 | −.78 |
| Random assignment | 343 | 3 | 0.81 | 0.93 | 1.71 | 1.3 | −1.26 |
| Nonrandom assignment | 180 | 5 | 1.30 | 1.49 | 0.78 | 23.8 | .49 |
| Studies without control groups | 472 | 6 | 1.12 | 1.28 | 1.10 | 2.1 | −.13 |
| Procedural knowledge–skills (all studies) | 2,109 | 32 | 1.09 | 1.18 | 1.18 | 4.0 | −.33 |
| Studies with control groups | 1,504 | 22 | 1.09 | 1.18 | 1.17 | 5.4 | −.20 |
| Random assignment | 1,095 | 15 | 0.90 | 0.97 | 0.63 | 11.8 | .17 |
| Nonrandom assignment | 253 | 5 | 0.60 | 0.65 | 0.00 | 100.0 | .65 |
| Studies without control groups | 605 | 10 | 1.09 | 1.19 | 1.38 | 1.9 | −.58 |
| Attitudes (all studies) | 1,929 | 52 | 0.29 | 0.33 | 0.17 | 56.9 | .11 |
| Studies with control groups | 694 | 12 | 0.32 | 0.36 | 0.21 | 47.0 | .10 |
| Random assignment | 455 | 6 | 0.32 | 0.36 | 0.18 | 58.8 | .14 |
| Nonrandom assignment | 212 | 5 | 0.37 | 0.42 | 0.24 | 32.1 | .11 |
| Studies without control groups | 1,235 | 40 | 0.27 | 0.30 | 0.14 | 65.7 | .12 |
| Job behavior | 2,513 | 66 | 0.25 | 0.27 | 0.31 | 22.6 | −.12 |
| Studies with control groups | 1,176 | 16 | 0.19 | 0.21 | 0.26 | 32.6 | −.13 |
| Random assignment | 581 | 7 | 0.14 | 0.15 | 0.38 | 16.6 | −.34 |
| Nonrandom assignment | 525 | 6 | 0.22 | 0.24 | 0.00 | 100.0 | .24 |
| Studies without control groups | 1,337 | 50 | 0.30 | 0.33 | 0.28 | 22.3 | −.03 |
| Results: workgroup productivity | 793 | 33 | 0.12 | 0.13 | 0.00 | 100.0 | .13 |
| Results: workgroup climate | 1,202 | 36 | 0.10 | 0.11 | 0.41 | 23.7 | −.41 |
| Studies with control groups | 335 | 3 | 0.07 | 0.08 | 0.00 | 100.0 | .08 |
| Studies without control groups | 867 | 33 | 0.11 | 0.13 | 0.47 | 23.3 | −.47 |

*Note.* Effect sizes for job behavior, workgroup productivity, and workgroup climate were all from supervisory training. Missing breakdowns are the result of too few studies ($k < 3$) for at least one of the comparisons. Studies that could not be unambiguously assigned to a breakdown condition (e.g., random assignment vs. nonrandom assignment) but which clearly fit in the higher level set (e.g., studies with control groups) were included in the higher level set but not the lower level sets, and thus sums of $N$ and $k$ for subsets may be less than $N$ and $k$ for the associated total set. Mean $d$ = sample-weighted mean of observed (uncorrected) effect sizes; $\delta$ = the population effect size estimate; % variance accounted for = percentage of variance in effect sizes attributable to sampling error.

substantial variability even after sampling error was removed. Estimates for BMT effects on results (workgroup productivity and climate) are smaller still ($\delta = 0.13$ and 0.11, respectively).

Turning now to methodological variables, studies without control groups, having been corrected for control group effects, yielded effect size estimates that are quite similar to those for studies with control groups. Smaller effects were observed for those control group studies using random assignment of participants to condition than for studies lacking random assignment for all outcomes other than procedural knowledge–skills. As indicated in Table 3, effect sizes were somewhat larger for published than unpublished studies across training outcomes (with the exception of declarative knowledge). We found no evidence of a positive bias in effects of studies provided by training firms.

## *Skills Taught*

Table 4 reports breakdowns of BMT effect size estimates based on the skills taught: interpersonal (either supervisory or teamwork skills) and technical skills. Training effects on the development of procedural knowledge–skills for technical skills training were slightly larger than for interpersonal skills training, although smaller for changes in trainees' attitudes. Studies reporting the use of BMT to train technical skills, such as use of computers or software, have generally been limited to assessing learning outcomes; hence, all results we report for behavior and results outcomes are limited to BMT applied to teaching interpersonal skills.

Supervisory skills training applications yielded larger average effects on the development of both declarative and procedural knowledge–skills than did teamwork skills training, whereas the latter was associated with larger training effects for changes in both attitudes and job behavior. Such differences may be the result of the content of these different types of programs. It must, however, be recognized that, in contrast to supervisory skills training, participants in the teamwork skills training studies were almost all nonsupervisory staff. Perhaps BMT is more effective in changing nonsupervisory employees' job behavior than that of supervisory-level staff, while being more effective in leading to changes in knowledge–skills for supervisory staff.

## *Stability of BMT Effects Over Time*

In addition to determining the overall effects of BMT on various outcomes, we were interested in determining whether learning and behavior change are maintained over time. In a small number of studies measuring learning outcomes, two posttests were taken on the same dependent variable at different times, one immediately after training and the other some months later, permitting a within-studies comparison of training effects across the two posttests. Table 5 provides the results of such an analysis for declarative knowledge and procedural knowledge–skills, the only two training outcomes with at least three studies including both immediate and delayed posttests. Effect sizes differences for declarative knowledge suggest an average decay of 0.24 standard deviations, whereas no decay was apparent for procedural knowledge–skills.

Another approach to assessing the extent to which BMT effects are maintained over time is to consider the relationship between study effect sizes and the length of time between training and when posttests were administered. To accomplish this, we regressed study training effect sizes on the number of months between training and posttest (which ranged from 0–12 months), using WLS regression in which studies were weighted by sample size. For studies that included both immediate and delayed posttests, we used the delayed posttest effect size in order to maximize variance in length of time between training and posttest for this analysis because, in most studies, posttests were administered shortly after training.

Table 3
*Breakdown of Behavior Modeling Training Effects on Training Outcomes by Publication Source*

| Subset of studies | *N* | No. of studies | Mean *d* | $\delta$ | Corrected *SD* of $\delta$ | % variance accounted for | Lower bound 90% credibility value |
|---|---|---|---|---|---|---|---|
| Declarative knowledge | | | | | | | |
| Published studies | 413 | 6 | 0.69 | 0.79 | 0.72 | 9.6 | −.12 |
| Unpublished studies | 582 | 8 | 1.30 | 1.49 | 1.59 | 2.7 | −.55 |
| Procedural knowledge–skills | | | | | | | |
| Published studies | 696 | 11 | 1.21 | 1.32 | 1.42 | 3.7 | −.50 |
| Unpublished studies | 1,373 | 20 | 0.98 | 1.06 | 0.99 | 5.2 | −.20 |
| From training firms | 245 | 3 | 0.48 | 0.53 | 0.00 | 100.0 | .53 |
| Not from training firms | 1,128 | 17 | 1.09 | 1.18 | 1.05 | 4.8 | −.17 |
| Attitudes | | | | | | | |
| Published studies | 184 | 4 | 0.56 | 0.64 | 0.00 | 100.0 | .64 |
| Unpublished studies | 1,745 | 48 | 0.26 | 0.29 | 0.15 | 62.5 | .10 |
| From training firms | 1,095 | 38 | 0.25 | 0.29 | 0.14 | 68.0 | .11 |
| Not from training firms | 650 | 10 | 0.27 | 0.30 | 0.17 | 52.8 | .09 |
| Job behavior | | | | | | | |
| Published studies | 573 | 9 | 0.33 | 0.35 | 0.26 | 38.0 | .02 |
| Unpublished studies | 1,940 | 57 | 0.27 | 0.29 | 0.32 | 19.4 | −.12 |
| From training firms | 1,256 | 48 | 0.30 | 0.33 | 0.24 | 31.5 | .02 |
| Not from training firms | 684 | 9 | 0.20 | 0.21 | 0.40 | 11.7 | −.30 |

*Note.* Missing breakdowns are the result of too few studies ($k < 3$) for at least one of the comparisons. Mean *d* = sample-weighted mean of observed (uncorrected) effect sizes; $\delta$ = the population effect size estimate; % variance accounted for = percentage of variance in (corrected) effect sizes attributable to sampling error.

Table 4
*Breakdown of Behavior Modeling Training Effects on Training Outcomes by Skills Taught*

| Subset of studies | *N* | No. of studies | Mean *d* | δ | Corrected *SD* of δ | % variance accounted for | Lower bound 90% credibility value |
|---|---|---|---|---|---|---|---|
| Declarative knowledge (interpersonal communication skills) | | | | | | | |
| Supervisory skills | 271 | 4 | 2.04 | 2.34 | 1.18 | 5.8 | .83 |
| Teamwork skills | 157 | 4 | 1.29 | 1.48 | 0.86 | 18.9 | .37 |
| Procedural knowledge–skills | | | | | | | |
| Interpersonal communication skills | 1,691 | 27 | 1.06 | 1.15 | 1.11 | 4.6 | −.26 |
| Supervisory skills | 1,152 | 17 | 1.27 | 1.38 | 1.24 | 3.4 | −.21 |
| Teamwork skills | 87 | 4 | 0.91 | 0.98 | 0.00 | 100.0 | .98 |
| Technical skills | 363 | 3 | 1.29 | 1.39 | 1.51 | 2.6 | −.54 |
| Attitudes | | | | | | | |
| Interpersonal communication skills | 1,300 | 46 | 0.31 | 0.36 | 0.16 | 65.2 | .15 |
| Supervisory skills | 1,042 | 39 | 0.28 | 0.32 | 0.14 | 73.2 | .15 |
| Teamwork skills | 144 | 4 | 0.51 | 0.58 | 0.00 | 100.0 | .58 |
| Technical skills | 597 | 5 | 0.19 | 0.22 | 0.00 | 100.0 | .22 |
| Job behavior (interpersonal communication skills) | | | | | | | |
| Supervisory skills | 2,073 | 57 | 0.26 | 0.28 | 0.35 | 20.7 | −.16 |
| Teamwork skills | 310 | 7 | 0.35 | 0.37 | 0.17 | 33.0 | .16 |

*Note.* *Teamwork skills* refers to peer-to-peer communication skills. Missing breakdowns are the result of too few studies ($k < 3$) for at least one of the comparisons. For procedural knowledge–skills and attitudes, *N*s and *k*s for interpersonal skills studies are larger than the sums of supervisory and teamwork skills studies because too few studies of other interpersonal skills were available for subset analyses. Mean *d* = sample-weighted mean of observed (uncorrected) effect sizes; δ = the population effect size estimate; % variance accounted for = percentage of variance in (corrected) effect sizes attributable to sampling error.

The results of these WLS regression analyses are reported in Table 6. Consistent with the results of the within-studies analyses, we found a small, negative relationship between time between training and posttest for declarative knowledge criteria. Surprisingly, we found a moderately large, positive (and statistically significant) relationship between training effects for procedural knowledge–skills and length of training–posttest interval, and we found a small positive (nonsignificant) relationship for job behavior. Taken together, these within-studies and between-studies analyses suggest that, whereas declarative knowledge appears to decay after training, newly learned skills (i.e., measured with tests of procedural knowledge or ratings of job behavior) are maintained or even increased over time after training.

### *Characteristics of BMT Design*

Results for the evaluation of various features of BMT design are presented in Tables 7–11. Where sufficient numbers of primary studies have experimentally contrasted a characteristic of training design, we report meta-analytic summary statistics for these within-studies comparisons, and where sufficient numbers of primary studies have applied a particular design characteristic and other studies have not, we report summary statistics in the form of between-studies breakdowns. When analyzing the effect on training outcomes of hours of training (see Table 11), a continuous moderator variable, we report results of WLS regressions.

These results generally support BMT design options that have been advocated in the literature, with a few notable exceptions. As expected, the development of procedural knowledge–skills is enhanced when learning points are used and when they are presented as rule codes (see Table 7). However, use of retention aids (e.g., small cards listing learning points) and the displaying of learning points along with modeling not only failed to contribute to greater development of knowledge or skills, but they may actually have hindered learning, as training effects on both declarative knowledge and skill development were smaller when retention aids were used and learning points were displayed during modeling.

Mixed (negative and positive) models, as compared with only positive models, yielded smaller training effects for the development of declarative knowledge and change in attitudes. In contrast, such mixed models showed greater transfer of training (changes in

Table 5
*Within-Study Comparison of Behavior Modeling Training Effects Based on Immediate Versus Delayed Posttests*

| Subset of studies | *N* | No. of studies | Sample-weighted mean of $(d_{\text{immediate}} - d_{\text{delayed}})$ | *SD* of $(d_{\text{immediate}} - d_{\text{delayed}})$ |
|---|---|---|---|---|
| Declarative knowledge | 284 | 3 | .24 | .83 |
| Procedural knowledge–skills | 307 | 3 | −.08 | .18 |

*Note.* A positive sample-size weighted mean difference between *d* values indicates that, on average, the effect size calculated using the immediate posttest was larger than the effect size calculated using the delayed posttest within the same study.

Table 6
*Relationship Between (Uncorrected) Training Effect Size and Number of Months Between Training and Posttest*

| Training outcome | *N* of studies | $\beta$ |
|---|---|---|
| Declarative knowledge | 11 | −.14 |
| Procedural knowledge–skills | 27 | .47* |
| Attitudes | 14 | .00 |
| Job behavior | 16 | .17 |

*Note.* Effect sizes were regressed on the number of months between training and posttest (ranging from 0–12), using weighted least-squares regression, in which studies were weighted by sample size. Training outcomes for which data from at least 10 studies were available are reported. A positive beta value indicates that larger effect sizes are associated with longer periods between training and posttest.
* $p < .05$.

job behavior) than did only positive models (see Table 8). The effect of models on the development of procedural knowledge–skills was less clear, with mixed models showing an increase of one half of a standard deviation in skill development over positive-only models in the within-studies comparison, but virtually no difference appeared in the between-studies comparison.

Symbolic rehearsal of skills before the behavioral rehearsal was associated with higher levels of skill development (see Table 9), but studies that included the verbal coaching of trainees prior to the behavioral rehearsal yielded slightly smaller training effects on skill development than those studies that did not mention use of coaching. Transfer of training (i.e., training effect on job behavior) was greatest when at least some of the scenarios that trainees practiced were trainee generated (see Table 9), when trainees were instructed to set goals, when trainees' managers were also trained, and when rewards and sanctions were instituted in trainees' workplaces for, respectively, using or not using newly learned skills (see Table 10).

We explored whether the number of hours of training—a proxy for the amount of time trainees practiced—would be positively associated with learning and transfer outcomes by regressing study effect sizes on the number of hours of training, using WLS regression. Table 11 presents standardized betas representing the relationship between training effect sizes and hours of training for various training outcomes. Consistent with the theory that more practice leads to greater skill development, the beta representing the relationship between hours of training and the development of procedural knowledge–skills (0.42) is positive, moderate in size, and approached significance. This advantage for longer training courses, however, does not seem to transfer to increased use of trained skills on the job, as reflected in the near-zero relationship between hours of training and job behavior.

## Discussion

The present study updates earlier research on BMT effects with a substantially larger pool of both published and unpublished

Table 7
*Behavior Modeling Training Effects on Learning Outcomes by Characteristics of Learning Points*

| Subset of studies | *N* | No. of studies | Mean *d* | $\delta$ | Corrected *SD* of $\delta$ | % variance accounted for | Lower bound 90% credibility value |
|---|---|---|---|---|---|---|---|
| Presentation of learning points (vs. no learning points) | | | | | | | |
| Procedural knowledge–skills: within-studies comparison | 240 | 4 | 0.76 | 0.82 | 0.11 | 86.6 | .69 |
| Presentation of learning points as rule codes (vs. as descriptions/summaries of behavior) | | | | | | | |
| Procedural knowledge–skills: within-studies comparison | 140 | 3 | 0.72 | 0.78 | 0.00 | 100.0 | .78 |
| Use of retention aids | | | | | | | |
| Declarative knowledge | | | | | | | |
| Retention aids provided to trainees | 708 | 7 | 0.73 | 0.84 | 1.17 | 1.8 | −.66 |
| No mention of use of retention aids | 287 | 7 | 1.82 | 2.08 | 1.39 | 8.5 | .31 |
| Procedural knowledge–skills | | | | | | | |
| Retention aids provided to trainees | 1,406 | 18 | 1.01 | 1.10 | 1.04 | 3.9 | −.24 |
| No mention of use of retention aids | 703 | 14 | 1.25 | 1.35 | 1.41 | 4.3 | −.45 |
| Display of learning points during modeling | | | | | | | |
| Declarative knowledge | | | | | | | |
| Learning points displayed with model | 777 | 8 | 1.00 | 1.15 | 1.46 | 1.4 | −.72 |
| No mention of learning points being displayed with model | 218 | 6 | 1.20 | 1.38 | 0.74 | 22.8 | .43 |
| Procedural knowledge–skills | | | | | | | |
| Learning points displayed with model | 1,522 | 20 | 0.95 | 1.03 | 1.03 | 4.0 | −.30 |
| No mention of learning points being displayed with model | 587 | 12 | 1.46 | 1.58 | 1.44 | 4.6 | −.26 |

*Note.* Within-studies comparisons are of experimental studies, in which a focal condition (e.g., presentation of learning points) was contrasted with another condition (e.g., no learning points): A positive mean effect reflects better performance, on average, for the focal condition. Mean $d$ = sample-weighted mean of observed (uncorrected) effect sizes; $\delta$ = the population effect size estimate; % variance accounted for = percentage of variance in (corrected) effect sizes attributable to sampling error.

Table 8
*Behavior Modeling Training Effects on Training Outcomes by Characteristics of the Modeling Display*

| Subset of studies | *N* | No. of studies | Mean *d* | δ | Corrected *SD* of δ | % variance accounted for | Lower bound 90% credibility value |
|---|---|---|---|---|---|---|---|
| Declarative knowledge | | | | | | | |
| Mixed models | 640 | 6 | 0.71 | 0.82 | 1.22 | 1.7 | −.75 |
| Positive-only models | 286 | 7 | 1.13 | 1.30 | 0.65 | 23.6 | .47 |
| Procedural knowledge–skills | | | | | | | |
| Mixed models | 1,123 | 15 | 1.12 | 1.22 | 1.13 | 3.4 | −.23 |
| Positive-only models | 917 | 16 | 1.14 | 1.24 | 1.24 | 4.6 | −.34 |
| Within-study comparison | 194 | 4 | 0.49 | 0.53 | 0.28 | 57.9 | .17 |
| Attitudes | | | | | | | |
| Mixed models | 1,509 | 46 | 0.27 | 0.30 | 0.15 | 64.3 | .11 |
| Positive-only models | 354 | 5 | 0.34 | 0.39 | 0.22 | 37.0 | .10 |
| Job behavior | | | | | | | |
| Mixed models | 1,875 | 58 | 0.32 | 0.34 | 0.30 | 22.3 | −.04 |
| Positive-only models | 548 | 7 | 0.13 | 0.14 | 0.25 | 38.7 | −.18 |

*Note.* The within-study comparison was of experimental studies, in which mixed model conditions were contrasted with positive-only model conditions. The positive mean effect for the within-study comparison reflects better performance, on average, for the mixed-model conditions. Mean *d* = sample-weighted mean of observed (uncorrected) effect sizes; δ = the population effect size estimate; % variance accounted for = percentage of variance in (corrected) effect sizes attributable to sampling error.

studies. It provides a similar, though slightly revised, picture of the effects of BMT in organizations. Whereas BMT effects on knowledge–skill (learning) outcomes were similar to those reported by Burke and Day (1986)—approximately 1 standard deviation improvement due to BMT—the effects of BMT on job behavior (all studies δ = 0.27) were considerably smaller than Burke and Day's previous estimate of 0.70. BMT effects on results outcomes were smaller still (δ = 0.11 for workgroup productivity and 0.13 for workgroup climate), and these are smaller than training effects on results that have been reported in earlier meta-analyses for training programs (Arthur, Bennett, Edens, & Bell, 2003; Burke & Day, 1986; Guzzo, Jette, & Katzell, 1985), which have ranged from 0.62 to 0.78.

We suspect that the smaller effects we found for BMT on job behavior and results are due to our (a) inclusion of published BMT studies that have become available since the early 1980s, some which have shown negligible effects of on transfer criteria (e.g., May & Kahnweiler, 2000; Russell et al., 1984; Werner et al., 1994), and (b) substantially larger proportions of unpublished studies than in these other meta-analyses. Previous meta-analyses on psychological interventions (e.g., Lipsey & Wilson, 1993) suggest that unpublished studies generally yield slightly smaller

Table 9
*Behavior Modeling Training Effects on Training Outcomes by Characteristics of the Behavioral Rehearsal*

| Subset of studies | *N* | No. of studies | Mean *d* | δ | Corrected *SD* of δ | % variance accounted for | Lower bound 90% credibility value |
|---|---|---|---|---|---|---|---|
| Trainees instructed to symbolically rehearse use of skills | | | | | | | |
| Procedural knowledge–skills: within-studies comparison | 313 | 5 | 0.27 | 0.39 | 0.27 | 51.6 | −.06 |
| Trainees coach each other prior to behavioral rehearsal | | | | | | | |
| Procedural knowledge–skills | | | | | | | |
| Coaching | 1,406 | 18 | 1.01 | 1.10 | 1.04 | 3.9 | −.24 |
| No mention of coaching | 656 | 13 | 1.30 | 1.41 | 1.45 | 4.3 | −.44 |
| Use of trainee-generated versus trainer-provided scenarios | | | | | | | |
| Job behavior | | | | | | | |
| Trainee-generated scenarios | 1,747 | 53 | 0.28 | 0.30 | 0.28 | 25.2 | −.06 |
| Trainer-provided scenarios | 109 | 3 | 0.14 | 0.16 | 0.12 | 48.0 | .00 |

*Note.* Within-studies comparisons are of experimental studies, in which a symbolic rehearsal condition was contrasted with a no symbolic rehearsal condition: The positive mean effect in the within-study comparison reflects better performance, on average, for the symbolic rehearsal condition. Mean *d* = sample-weighted mean of observed (uncorrected) effect sizes; δ = the population effect size estimate; % variance accounted for = percentage of variance in (corrected) effect sizes attributable to sampling error.

Table 10
*Behavior Modeling Training Effects on Job Behavior by Posttraining Transfer Enhancements*

| Subset of studies | $N$ | No. of studies | Mean $d$ | $\delta$ | Corrected $SD$ of $\delta$ | % variance accounted for | Lower bound 90% credibility value |
|---|---|---|---|---|---|---|---|
| Trainee instructed to set goals | | | | | | | |
| Goal-setting instructions included | 1,500 | 52 | 0.35 | 0.37 | 0.30 | 23.3 | −.01 |
| No mention of goal-setting instructions | 1,013 | 14 | 0.18 | 0.19 | 0.25 | 32.2 | −.13 |
| Trainees' managers also trained | | | | | | | |
| Managers also trained | 302 | 7 | 0.50 | 0.53 | 0.51 | 10.6 | −.12 |
| No mention of managers being trained | 2,211 | 59 | 0.27 | 0.29 | 0.30 | 23.5 | −.10 |
| Rewards/sanctions in the work environment | | | | | | | |
| Rewards/sanctions instituted | 223 | 4 | 0.51 | 0.55 | 0.00 | 100.0 | .55 |
| No mention of rewards/sanctions being instituted | 1,978 | 59 | 0.29 | 0.31 | 0.29 | 23.7 | −.07 |

*Note.* Missing breakdowns are the result of two few studies ($k < 3$) for at least one of the comparisons. Mean $d$ = sample-weighted mean of observed (uncorrected) effect sizes; $\delta$ = the population effect size estimate; % variance accounted for = percentage of variance in (corrected) effect sizes attributable to sampling error.

effect sizes than published studies as was the case in the present study for all outcomes other than declarative knowledge. An appropriate comparison of BMT and other training methods would be to compare the present results with meta-analyses of other training methods that have included a majority of unpublished studies. Morrow, Jarrett, and Rupinski (1997), for example, reported an average, uncorrected effect size of 0.31 for change in behavior resulting from 11 managerial training programs conducted within a large pharmaceutical firm, just slightly larger than that of the present study (uncorrected average $d$ of 0.25 for job behavior).

Finally, BMT effects on job behavior reported here should be considered a lower bound for what might have been achieved had these programs included recently developed posttraining interventions, such as self-management and relapse prevention. Studies that have demonstrated the superiority of adding these interventions to training (e.g., Gist, Stevens, & Bavetta, 1991; Stevens & Gist, 1997) were not included in this meta-analysis because they lacked no-training comparison conditions.

## *The Stability of BMT Effects Over Time*

Determining the extent to which training effects are stable (vs. decay) over time is critical to determining accurately the utility of training programs. If the immediate effects of training on the development of skills and changes in job behavior quickly dissipate, the utility of the program is severely reduced (Mathieu & Leonard, 1987). BMT differs from other training methods, in part, due to its greater emphasis on the transfer of skills to the workplace and maintenance of those skills over time (Decker & Nathan, 1985; Goldstein & Sorcher, 1974).

Table 11
*Relationship Between (Uncorrected) Training Effect Size and Hours of Training*

| Training outcome | $N$ of studies | $\beta$ |
|---|---|---|
| Procedural knowledge–skills | 21 | .42[a] |
| Attitudes | 22 | −.12 |
| Job behavior | 28 | −.01 |
| Results: workgroup productivity | 12 | .28 |
| Results: workgroup climate | 14 | .12 |

*Note.* Effect sizes were regressed on the number of hours of training using weighted least-squares regression, in which studies were weighted by sample size. Training outcomes for which data from at least 10 studies were available are reported. A positive beta value indicates that larger effect sizes are associated with more training.
[a] This value approached significance, $p < .06$.

Results from both within-studies comparisons of delayed versus immediate posttests and the regression of BMT effect sizes on months between training and posttests suggest that declarative knowledge decays over time. This finding is consistent with the typical curves characterizing recognition and recall over time, due to decay (Brown, 1958; Peterson & Peterson, 1959), interference (Postman & Underwood, 1973), or retrieval failure (Tulving & Thompson, 1971, 1973).

In contrast, however, skills and changed job behavior were clearly sustained over time after BMT training. When comparing training effect sizes based on immediate and delayed posttests within studies, we found that effect sizes for procedural knowledge–skills derived from delayed posttests were, on average, slightly larger than those derived from posttests taken immediately after training. Furthermore, when regressing effect sizes on the number of months between training and posttest, we found *positive* relationships for measures of both procedural knowledge–skills ($\beta = .47$, $p < .05$) and on-the-job behavior ($\beta = .17$, *ns*), again suggesting that changes in skills, and to a lesser extent job behavior, are actually greater with the passage of time after training.

The large relationship found between BMT effects on skills and length of time between training and posttest is somewhat surprising in light of both the assumption typically made in training utility analyses that training effects decay over time (see, e.g., Cascio, 1989; Mathieu & Leonard, 1987; Wexley & Latham, 2002) and consistent evidence of skill decay over posttraining periods of no practice (Arthur et al., 1998). We suspect that trainees' skills may continue to improve after BMT because of the opportunities for practice that naturally occur on the job. Such practice opportunities are likely to be valuable for skill development but may not occur frequently enough for trainees to have sufficient practice opportunities within just the 1st month following training (May & Kahnweiler, 2000), when posttraining evaluation measures are often taken.

Although the positive relationships between effect sizes and months between training and posttest for procedural knowledge–skills and job behavior are encouraging, two study limitations have to be recognized. First, because of the correlational nature of such analyses, we are unable to rule out third-variable causes, such as the possibility of more rigorous implementations of training, follow-up activities, or special incentives provided to trainees in studies using delayed posttests. Second, the time between training and posttests failed to exceed 1 year in all studies, and in most studies it was 6 months or less, and so the extrapolation of these findings over subsequent years following training must be done cautiously. Clearly, there is a need for more primary studies that assess training effects for periods beyond 1 year.

### *Characteristics of BMT Design*

A number of features of BMT design were associated with larger effects on learning and changes in job behavior. We discuss the implications of each of these findings next.

*Learning points.* We found that presenting learning points, and specifically the presentation of learning points in the form of rule codes, yields large effects (.82 and .78, respectively) on trainees' development of procedural knowledge–skills. These findings provide a quantitative estimate of effects summarized earlier by Decker and Nathan (1985). We were unable to find sufficient numbers of studies in which learning points were either not presented or not presented as rule codes for us to be able to conduct a between-studies analysis of learning points (i.e., we had to rely entirely on studies in which two or more learning point conditions were experimentally compared), which speaks to how effectively previous research concerning learning points has diffused into practice.

We also found that retention aids and displays of learning points, along with the behavior models, were associated with smaller gains in declarative knowledge and skill development through training. Such results may be related to issues of multitasking, in which trainees may attend to these tasks sequentially, may use some sort of filtering mechanism, or may alternate between different sources of information (Kahneman, 1973; Lindsay & Norman, 1972; Wickens & Seidler, 1997), resulting in interference with learning.

*Positive-only versus mixed models.* Transfer of training (i.e., changes in job behavior) was greatest when mixed (negative and positive) models, rather than only positive models, were presented to trainees. Baldwin (1992) demonstrated that mixed models performed better in a task requiring the generalization of skills to a simulation different from that which was modeled. The present results support those findings and extend them to the transfer of skills to job settings and to nonstudent samples.

Baldwin (1992) suggested that contrasting effective and ineffective behaviors within the modeling display is likely to provide greater distinctiveness of the stimuli. Furthermore, mixed models may lead to greater motivation when modeling displays depict the outcomes of the model's behavior. If mixed-model displays show that the model using effective behaviors achieves his or her desired outcome (i.e., those behaviors are reinforced), whereas ineffective behaviors fail to achieve outcomes, trainees' motivation to use effective behaviors and to avoid using ineffective behaviors may be greater than if they view only effective behaviors achieving desired outcomes.

Not surprisingly, mixed models were not superior to positive-only models in producing changes in trainees' attitudes, and they were inferior to positive-only models in improving trainees' declarative knowledge. The inclusion of negative models may actually distract trainees from remembering declarative information, such as the names of rule codes. Similarly, Baldwin (1992) found that trainees' reproduction of skills in a simulation identical to that displayed in a modeling video was superior in a positive-only condition.

In the case of teaching technical skills, where there may be only single behavioral options for achieving particular tasks, the inclusion of negative models may not be appropriate. In these cases, simple reproduction of the model's behavior is likely to be important and may be best achieved through positive-only modeling (Baldwin, 1992). If, however, a variety of responses can be used when facing a target task, as is the case in interpersonal tasks or more complex technical skills, the mixed model may aid in generalizing the response to the transfer situation. Furthermore, if trainees are likely to have developed habitual responses (that are less than ideally effective) through having faced target tasks prior to training, mixed models may lead to greater skill development and transfer because trainees unlearn ineffective responses through viewing models who use a repertoire of responses, including those that are similar to the trainees'. Virtually all of the studies contributing to our analyses of mixed and positive-only models were of BMT applied to teaching interpersonal skills, prohibiting a breakdown of display model characteristics by skills taught. With the recent expansion of BMT to a broader range of skills, further research concerning the efficacy of modeling displays for various types of tasks is needed.

*Symbolic and behavioral rehearsals.* Asking trainees to rehearse symbolically (mentally) just before the behavioral rehearsal component of BMT resulted in a small (0.39 standard deviations) average increase in procedural knowledge–skills demonstrated during the behavioral rehearsal. However, coaching trainees in how they will apply newly learned skills during the behavioral rehearsal—a similar, though more overt, preparation strategy—was not associated with greater procedural knowledge–skills. We suspect that these different outcomes for quite similar preparation strategies may have resulted from a methodological artifact: Unlike studies of symbolic rehearsal, in which procedural knowledge–skills were measured during the behavioral rehearsal, the studies for which we compared the presence–absence of coaching measured procedural knowledge–skills after training, and so initial effects of coaching may have become defused as trainees participated in behavioral rehearsals prior to testing.

In the case of both symbolic rehearsal and coaching, the effects of these additions to BMT may be strongest for procedural knowledge–skills demonstrated during the initial behavioral rehearsal component of training. Unfortunately, too few studies measuring transfer outcomes reported whether symbolic rehearsal or coaching were used, and so an assessment of posttraining effects of either was not possible. Further research as to the transfer effects of symbolic rehearsal and coaching is needed.

Training effects on changes in job behavior were greater when, during behavioral rehearsals, trainees practiced work-related scenarios that they developed themselves, as opposed to practicing only scenarios provided by trainers. This finding is consistent with both the principle of identical elements and encoding specificity theory (Tulving & Thompson, 1971, 1973) and also with the more recent theory of situated cognition (e.g., Anderson, Reder, & Simon, 1996; Greeno, 1998), as trainees developing and practicing their own work-related scenarios may have resulted in more authentic learning experiences. Alternatively, the use of trainee-generated practice scenarios may lead to greater transfer by facilitating greater retention or self-efficacy for training transfer. Having trainees generate work-related practice scenarios may lead to greater retention through the additional cognitive elaboration that is required. Greater self-efficacy for transfer may be achieved through the social reinforcement trainees receive from the trainer and other trainees after they have practiced and through successfully applying newly learned skills to work-related scenarios. Future research might clarify the mechanism(s) through which trainee-generated practice scenarios lead to increased training transfer.

*Hours of training.* The number of hours of training, a proxy measure of the amount of practice trainees received, was positively associated with larger training effects on the development of procedural knowledge–skills. However, we found a near-zero relationship between the number of hours of training and the extent of training transfer (i.e., size of training effects for changes in job behavior). The additional practice afforded in longer BMT programs may fail to increase training transfer because of lack of sufficient variability in practice conditions. Additional practice in training that repeats the same task, feedback conditions, and context can result in greater skill acquisition at the expense of generalization of skills to the transfer setting because attentional demands of trainees are not sufficiently maintained and because stimuli and required responses in the training setting fail to match the full range of situations that will be faced on the job (Arthur, Day, Bennett, McNelly, & Jordan, 1997; Campbell & Kuncel, 2001; Ghodsian, Bjork, & Benjamin, 1997; Goettl, Yadrick, Connolly-Gomez, Regian, & Shebilske, 1996; R. A. Schmidt & Bjork, 1992; Shute & Gawlick, 1995). BMT in organizational settings has typically been structured in a modular format, so that longer training programs simply include additional modules that teach very similar behaviors (i.e., repeating many of the same learning points) under the same modeling and practice conditions. Such programs have been criticized for too much repetition, which leads to boredom among trainees (Parry & Reich, 1984). The lack of evidence for a relationship between additional training (and practice) and improved transfer found in the present study implies that the incremental costs of providing lengthy BMT programs, in which practice sessions are all very similar, may not be warranted.

*Posttraining transfer enhancers.* We explored three specific strategies that have been hypothesized to lead to greater transfer of BMT: having trainees set goals as to how they will apply newly learned skills, the training of trainees' superiors, and the institution of rewards and sanctions in trainees' work environments for, respectively, use or lack of use of newly learned skills.

Training effect size estimates for changes in job behavior were slightly larger for studies applying goal setting than for studies without goal setting. Goal setting has been found to be an effective posttraining strategy (Werner et al., 1994; Wexley & Baldwin, 1986; Wexley & Nemeroff, 1975), and the lack of more pronounced differences in effect sizes based on goal setting in the present study may have resulted from weak or varied implementations of goal setting in our primary studies (e.g., trainees may not have been encouraged to set specific goals, and posttraining follow-up/feedback on goals may have been lacking). Furthermore, even within studies in which trainees were not encouraged to set goals, some trainees may still have initiated goal setting themselves. Similarly, no study explicitly stated that trainees were not encouraged by trainers to set goals, and so it is possible that some of the studies coded as not including goal setting in fact did so, without mention of this in their report.

As expected, the training of trainees' superiors was strongly associated with larger training effect sizes for changes in job behavior of trainees. Effect size was approximately 0.30 larger for studies in which trainees' superiors were also trained. The training of trainees' superiors is likely to affect motivational processes in a variety of ways. Superiors who have attended the training are in a stronger position to provide managerial support for training (Baldwin & Ford, 1988; Huczynski & Lewis, 1980), such as modeling skills in the workplace, discussing with trainees how newly learned skills can be applied to the work environment (Brinkerhoff & Montesino, 1995), and providing social reinforcement when trainees use skills on the job (Holton, Bates, & Ruona, 2000; Holton, Bates, Seyler, & Carvalho, 1997).

Similarly, studies that reported the introduction of rewards/sanctions for trainees' using/not using newly learned skills yielded substantially larger BMT effect size estimates for job behavior. Examples of such rewards/sanctions reported in primary studies included integration of newly learned skills into performance appraisals, encouraging trainees' superiors to direct and reinforce their subordinates in using newly learned skills. These findings are consistent with a growing body of research on the influence of work environment variables on transfer of training (for reviews, see Ford & Weissbein, 1997; Salas & Cannon-Bowers, 2001).

## Limitations and Conclusions

As with most meta-analyses, missing information was encountered in many original studies, leading to the exclusion of some studies from the present analysis. Missing data also precluded corrections for all theoretical aspects of measurement error and prevented further analyses, for example, refinements on the presentation of learning points (e.g., trainee-generated learning points), the modeling component (e.g., the inclusion of learning points in modeling videos, having trainees take notes during modeling, use of multiple rather than single models, mixed vs. positive-only models for technical training), and behavioral rehearsal (e.g., spaced vs. massed practice, skill practice group size, use of video feedback). Insufficient information was most prevalent in unpublished reports provided by training firms. Although training firms' encouragement of their clients to collect evaluation information on training programs is laudable, more careful documentation of the organizational context in which the training was conducted, as well as basic descriptive statistics (group means and standard deviations; sample sizes; reliabilities of measures; and, in the case of repeated measures designs, pretest–posttest correlations), would be helpful.

For many training outcomes, considerable variance in training effects across studies remained after removing variance due to sampling error, as indicated by large residual variances and low lower bound 90% credibility values. Large residual variances often remained even after studies were broken down by methodological variables, suggesting that moderator variables other than those assessed in the present study are likely to be responsible for the remaining variability of BMT effects across studies. We suspect that some between-studies variability in BMT effects on learning outcomes may be due to differences in the sensitivity of measures used and organization-level differences in trainees' (pretraining) levels of skills and motivation (Colquitt, LePine, & Noe, 2000). In the case of job behavior and results (productivity and climate) outcomes, residual variability may result from unmeasured, organization-level differences in work environment variables other than those assessed in the present study.

Training practitioners may be tempted to calculate the return on investment of conducting BMT by applying training utility analysis (e.g., Cascio, 1989) and using the population effect size estimates for job behavior here, but doing so would result in an upwardly biased utility estimate. Measures of job behavior in virtually all studies included here were focused on specific aspects of trainees' performance that were expected to change as a result of training and did not assess trainees' performance across the entire performance domain of trainees' jobs. Consequently, these effect size estimates would overstate the effect that BMT would have on overall job performance to the extent of criterion deficiency in the performance measures used. The appropriate correction for utility analysis would involve either reducing obtained effect sizes by the extent of criterion deficiency or obtaining estimates of the standard deviation of job performance based on the dollar value of variation in only those aspects of performance measured in each study. Such information was not reported in the original studies included in this meta-analysis, and so utility estimates cannot be calculated directly.

Nevertheless, returns on investment with BMT training, based on effect size estimates reported here, are likely to be high enough to justify the value of using behavior modeling to improve performance in organizational settings. Morrow et al. (1997) collected necessary information to correct managerial training (not BMT) effect sizes in one organization for criterion deficiency and demonstrated an average return on investment of 45%, based on an average effect size of 0.31 (uncorrected). Mathieu and Leonard (1987) estimated even greater returns for a BMT program with an effect size estimate of 0.31 for overall job performance. Assuming comparability of other variables between the training contexts of studies included here and in both Morrow et al.'s meta-analysis and Mathieu and Leonard's study (e.g., extent of criterion deficiency, dollar value of performance, and training costs), the similar effect size found here for BMT could be expected to produce similar financial returns. The utility of BMT is further supported by the finding of no decay in training effects on skill use over time.

The body of published and unpublished evaluation research on BMT reviewed here has demonstrated the approach to be an effective, psychologically based training intervention that has been used to produce sustainable improvements in a diverse range of skills and posttraining behavior. We have highlighted features of training design that are associated with larger training effect sizes for learning and transfer outcomes, which can provide practical guidance for the application of BMT in the future.

## References

References marked with an asterisk indicate studies included in the meta-analysis.

*Alssid, L. L., & Hutchison, W. R. (1977). Comparison of modeling techniques in counselor training. *Counselor Education and Supervision, 17,* 36–41.

*American Telephone & Telegraph. (1974). *Analysis and evaluation of supervisory relationships training* (Internal report). New York: Author.

Anderson, J. R., Reder, L. M., & Simon, H. A. (1996). Situated learning and education. *Educational Researcher, 25*(4), 5–11.

*Andes, K. L., Ravishankar, L. P., & Russ-Eft, D. F. (1997). Transfer strategies supporting interpersonal skills training for supervisors: An evaluation study. In M. L. Broad (Ed.), *In action: Transferring learning to the workplace* (pp. 71–83). Washington, DC: American Society for Training and Development.

*Angelone, M. W. (1982). The application of behavior modeling to empathy training and its influence on managerial flexibility and level of empathy. *Dissertation Abstracts International, 43*(12), 3788–3789A. (UMI No. 8303134)

Arnold, J., & Davey, K. M. (1992). Self-ratings and supervisor ratings of graduate employees' competences during early career. *Journal of Occupation and Organizational Psychology, 65,* 235–250.

Arthur, W., Jr., Bennett, W. J., Edens, P. S., & Bell, S. T. (2003). Effectiveness of training in organizations: A meta-analysis of design and evaluation features. *Journal of Applied Psychology, 88,* 234–245.

Arthur, W., Jr., Bennett, W. J., Stanush, P. L., & McNelly, T. L. (1998). Factors that influence skill decay and retention: A quantitative review and analysis. *Human Performance, 11*(1), 57–101.

Arthur, W., Jr., Day, E. A., Bennett, W. J., McNelly, T. L., & Jordan, J. A. (1997). Dyadic versus individual training protocols: Loss and reacquisition of a complex skill. *Journal of Applied Psychology, 82,* 783–791.

*Baldwin, T. T. (1992). The effects of alternative modeling strategies on outcomes of interpersonal-skills training. *Journal of Applied Psychology, 77,* 147–154.

Baldwin, T. T., & Ford, J. K. (1988). Transfer of training: A review and directions for future research. *Personnel Psychology, 41,* 63–105.

Bandura, A. (1977). *Social learning theory.* Englewood Cliffs, NJ: Prentice Hall.

*Birkenbach, X. C., Kamfer, L., & ter Morshuizen, J. D. (1985). The development and the evaluation of a behavior-modeling training program for supervisors. *South African Journal of Psychology, 15,* 11–19.

*Blickensderfer, E., Cannon-Bowers, J. A., & Salas, E. (1997, April). *Training teams to self-correct: An empirical investigation.* Paper presented at the meeting of the Society for Industrial and Organizational Psychology, St. Louis, MO.

*Bostain, N. S. (2000). Evaluation of a management development behavior modeling training program in industry: Transfer of training. *Dissertation Abstracts International, 61*(4), 2252B. (UMI No. 9968772)

*Bright, J. E. H., & Halse, R. (1998). An evaluation of an industry training program. *Training and Development in Australia, 4,* 9–11.

Brinkerhoff, R. O., & Montesino, M. U. (1995). Partnership for training transfer: Lessons from a corporate study. *Human Resource Development Quarterly, 6,* 263–274.

*Brock, D. B. (1994). Computer training: The role of computer attitudes and behavior modeling in the acquisition of declarative and procedural knowledge. *Dissertation Abstracts International, 55*(3), 1212B. (UMI No. 9419869)

Brown, J. A. ( 1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology, 10,* 12–21.

*Bryant, S. E., & Fox, S. K. (1995). Behavior modeling training and generalization: Interaction of learning point type and number of modeling scenarios. *Psychological Record, 45,* 495–503.

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

*Burnaska, R. F. (1976). The effects of behavior modeling training on managers' behaviors and employees' perceptions. *Personnel Psychology, 29,* 329–335.

*Byham, W. C., Adams, D., & Kiggins, A. (1976). Transfer of modeling training to the job. *Personnel Psychology, 29,* 345–349.

Campbell, J. P., & Kuncel, N. R. (2001). Individual and team training. In N. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial, work and organizational psychology* (Vol. 1, pp. 278–312). Thousand Oaks, CA: Sage.

Carless, S. A., & Roberts-Thompson, G. P. (2001). Self-ratings in training programs: An examination of level of performance and the effects of feedback. *International Journal of Selection and Assessment, 9,* 217–225.

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,* 851–862.

Cascio, W. F. (1989). Using utility analysis to assess training outcomes. In I. L. Goldstein (Ed.), *Training and development in organizations* (pp. 63–88). San Francisco: Jossey-Bass.

*Catanzaro, D. (1993). The impact of an employee program on service quality in a nursing home organization. *Dissertation Abstracts International, 53*(8), 4403B. (UMI No. 9230204)

*Cellar, D. F., & Wade, K. (1988). Effect of behavioral modeling on intrinsic motivation and script-related recognition. *Journal of Applied Psychology, 73,* 181–192.

*Chou, H. W. (2001). Effects of training method and computer anxiety on learning performance and self-efficacy. *Computers in Human Behavior, 17,* 51–69.

Church, A. H. (1997). Do you see what I see? An exploration of congruence in ratings from multiple perspectives. *Journal of Applied Social Psychology, 27,* 983–1020.

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.

*Compeau, D. R., & Higgins, C. A. (1995). Application of social cognitive theory to training for computer skills. *Information Systems Research, 6,* 118–143.

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

Decker, P. J. (1980). Effects of symbolic coding and rehearsal in behavior-modeling training. *Journal of Applied Psychology, 65,* 627–634.

*Decker, P. J. (1982). The enhancement of behavior modeling training of supervisory skills by the inclusion of retention processes. *Personnel Psychology, 35,* 323–332.

*Decker, P. J. (1984). Effects of different symbolic coding stimuli in behavior modeling training. *Personnel Psychology, 37,* 711–720.

Decker, P. J., & Nathan, B. R. (1985). *Behavior modeling training: Principles and applications.* New York: Praeger.

*Development Dimensions International. (1986, July). *Assessment and development of administrative and interpersonal skills for MBA students at the University of Pittsburgh.* Pittsburgh, PA: Author.

Driskell, J. E., Copper, C., & Moran, A. (1994). Does mental practice enhance performance? *Journal of Applied Psychology, 79,* 481–492.

Driskell, J. E., Willis, R. P., & Copper, C. (1992). Effect of overlearning on retention. *Journal of Applied Psychology, 77,* 615–622.

*Dubois, J. (1994). *The perception of change in leadership competence, job satisfaction and organizational climates as a result of training by the behavior modeling method.* Unpublished master's thesis, Pacific Lutheran University.

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,* 170–177.

Falcone, A. J. (1985). Meta-analysis of personnel training techniques for three populations. *Dissertation Abstracts International, 47* (01), 412B. (UMI No. AAD86–06497)

Ford, J. K., & Weissbein, D. A. (1997). Transfer of training: An updated review and analysis. *Performance Improvement Quarterly, 10,* 22–41.

Furnham, A., & Stringfield, P. (1994). Congruence of self and subordinate ratings of managerial practices as a correlate of supervisor evaluation. *Journal of Occupational and Organizational Psychology, 67,* 57–67.

Ghodsian, D., Bjork, R., & Benjamin, A. (1997). Evaluating training during training: Obstacles and opportunities. In M. A. Quinones & A. Ehrenstein (Eds.), *Training for a rapidly changing workplace: Applications for psychological research* (pp. 63–88). Washington, DC: American Psychological Association.

*Gist, M., Rosen, B., & Schwoerer, C. (1988). The influence of training method and trainee age on the acquisition of computer skills. *Personnel Psychology, 41,* 255–265.

*Gist, M., Schwoerer, C., & Rosen, B. (1989). Effects of alternative training methods on self-efficacy and performance in computer software training. *Journal of Applied Psychology, 74,* 884–891.

Gist, M. E., Stevens, C. K., & Bavetta, A. G. (1991). Effects of self-efficacy and post-training intervention on the acquisition and maintenance of complex interpersonal skills. *Personnel Psychology, 44,* 837–861.

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

Goettl, B. P., Yadrick, R. M., Connolly-Gomez, C., Regian, W. J., & Shebilske, W. L. (1996). Alternating task modules in isochronal distributed training of complex tasks. *Human Factors, 38,* 330–346.

Goldstein, A. P., & Sorcher, M. A. (1974). *Changing supervisory behavior.* New York: Pergaman Press.

*Graham, S. L. (1984). The effects of behavior modeling and formal feedback on skills acquisition and attitude toward training. *Dissertation Abstracts International, 45*(10), 3118A. (UMI No. AAD84–25297)

Greeno, J. G. (1998). The situativity of knowing, learning, and research. *American Psychologist, 53,* 5–26.

*Grissom, C. M. (1986). Behavior modeling: The evaluation of a program to develop conflict management skills. *Dissertation Abstracts International, 47*(09), 3281A. (UMI No. AAD87–00713)

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.

Harris, M. M., & Schaubroeck, J. (1988). A meta-analysis of self–supervisor, self–peer, and peer–supervisor ratings. *Personnel Psychology, 41,* 43–62.

*Harrison, J. K. (1987). An examination of the individual and combined effects of behavior modeling and the cultural assimilator as cross-cultural management training techniques. *Dissertation Abstracts International, 48*(08), 2103A. (UMI No. AAD87–25398)

*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,* 952–962.

*Hauenstein, P. C. (1984). A comparative analysis of the underlying learning processes in behavior modeling: Social learning theory evaluated. *Dissertation Abstracts International, 44*(10), 3230B. (UMI No. 8402098)

*Hogan, P. M., Hakel, M. D., & Decker, P. J. (1986). Effects of trainee-generated versus trainer-provided rule codes on generalization in behavior-modeling training. *Journal of Applied Psychology, 71,* 469–473.

Holton, E. F., III., Bates, R., & Ruona, W. E. A. (2000). Development of a generalized learning transfer system inventory. *Human Resource Development Quarterly, 11,* 333–360.

Holton, E. F., III., Bates, R., Seyler, D. L., & Carvalho, M. B. (1997). Toward a construct validation of a transfer climate instrument. *Human Resource Development Quarterly, 8,* 95–113.

*Howard, C. K. (1992). *Training electronic equipment operators: Behavior modeling versus text versus trial and error.* Unpublished doctoral dissertation, Iowa State University, Ames.

Huczynski, A. A., & Lewis, J. W. (1980). An empirical study into the learning-transfer process in management training. *Journal of Management Studies, 17,* 227–240.

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.

Johnson, R. D., & Marakas, G. M. (2000). Research report: The role of behavioral modeling in computer skills acquisition: Toward refinement of the model. *Information Systems Research, 11,* 402–417.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice Hall.

*Katzenmeyer, M. H. (1988). An evaluation of behavior modeling training designed to improve selected skills of educational managers. *Dissertation Abstracts International, 49*(10), 2896A. (UMI No. AAD88–22453)

*King, H. D., & Arlinghaus, C. G. (1976). Interaction Management® validated in the steel industry. *Assessment and Development, 3*(2), 1–2.

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., & Saari, L. M. (1979). Application of social learning theory to training supervisors through behavioral modeling. *Journal of Applied Psychology, 64,* 239–246.

*Leonel, G. (1994). A study to determine the effectiveness of the Zenger-Miller behavior modeling frontline leadership program to improve interpersonal skills for paper industry supervisors. *Dissertation Abstracts International, 54*(12), 4376A. (UMI No. 9409998)

Lindsay, P. H., & Norman, D. A. (1972). *Human information processing: An introduction to psychology.* New York: Academic Press.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist, 48,* 1181–1209.

*Mann, R. B., & Decker, P. J. (1984). The effect of key behavior distinctiveness on generalization and recall in behavior modeling training. *Academy of Management Journal, 27,* 900–910.

*Marson, P. P. (1987). A study of behavior modeling in management training: A case study. *Dissertation Abstracts International, 49*(07), 1880A. (UMI No. AAD88–08077)

*Martens, C. A. (1992). *The effects of cognitive and behavior modeling on in-basket performance and self-efficacy beliefs.* Unpublished master's thesis, California State University, Long Beach.

*Martineau, J. W. (1996). A contextual examination of the effectiveness of a supervisory skills training program. *Dissertation Abstracts International, 56*(9), 5207B. (UMI No. 9600217)

*Mathieu, J. E., & Leonard, R. L. (1987). Applying utility concepts to a training program in supervisory skills: A time-based approach. *Academy of Management Journal, 30,* 316–335.

*Mattimore, K. L. (1993). A comprehensive evaluation of an interviewer skills workshop. *Dissertation Abstracts International, 54*(6), 3378B. (UMI No. 9319749)

*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,* 353–373.

Mayer, S. J., & Russell, J. S. (1987). Behavior modeling training in organizations: Concerns and conclusions. *Journal of Management, 13,* 21–40.

*McCullough, H. F. (1981). Role playing versus case studies in interpersonal skills training: Effects on the attitudes of middle level managers, their superiors, and subordinates. *Dissertation Abstracts International, 42*(08), 3729A. (UMI No. AAD82–01246)

McGhee, W., & Tullar, W. L. (1978). A note on evaluating behavior modification and behavior modeling as industrial training techniques. *Personnel Psychology, 31,* 477–484.

Meyer, H. H. (1980). Self appraisal of job performance. *Personnel Psychology, 33,* 291–295.

*Mills, G. E., & Pace, R. W. (1989). What effects do practice and video feedback have on the development of interpersonal communication skills? *The Journal of Business Communication, 26,* 159–176.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7,* 105–125.

Morrow, C. C., Jarrett, M. Q., & Rupinski, M. T. (1997). An investigation of the effect and economic utility of corporate-wide training. *Personnel Psychology, 50,* 91–119.

*Moses, J. L., & Ritchie, R. J. (1976). Supervisory relationships training: A behavioral evaluation of a behavior modeling program. *Personnel Psychology, 29,* 337–343.

*Moskal, P. J., Mulligan, C. P., & Carson, R. (1988). *Improving Navy recruit confidence expectancies and knowledge in a simulated chemical warfare environment* (NAVTRASYSCEN Publication No. TR88–010). Orlando, FL: Naval Training Systems Center.

*Mun, Y. Y. (1998). Developing and validating an observational learning model of computer software training. *Dissertation Abstracts International, 59*(10), 3889A. (UMI No. 9909036)

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

*Newman, J. L., & Fuqua, D. R. (1988). A comparative study of positive and negative modeling in counselor training. *Counselor Education and Supervision, 28,* 121–129.

*Nunns, C. G., & Bluen, S. D. (1992). The impact of behavior modeling training on self-reports of White supervisors in two South African mines. *Journal of Applied Behavioral Science, 28,* 433–444.

*Orpen, C. (1985). The effects of behaviour modelling training on managerial attitudes and performance: A field experiment. *International Journal of Manpower, 6,* 21–24.

*Palmer, S. N. (1981). A field study of the dissonance hypothesis in a behavior modeling training program. *Dissertation Abstracts International, 42*(07), 3079A. (UMI No. AAD81–28542)

Parry, S. B., & Reich, L. R. (1984). An uneasy look at behavior modeling. *Training and Development Journal, 38*(3), 57–62.

Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58,* 193–198.

Postman, L., & Underwood, B. J. (1973). Critical issues in interference theory. *Memory & Cognition, 1,* 19–40.

Renaud, L., & Stolovitch, H. (1988). Simulation gaming: An effective strategy for creating appropriate traffic safety behaviors in five-year-old children. *Simulation & Games, 19,* 328–345.

*Robertson, I. T., Bell, R., & Sadri, G. (1991). Behavior modeling training: Variations in retention processes. *Personnel Review, 20*(4), 25–28.

*Robinson, J. C. (1982). *Developing managers through behavior modeling.* San Diego, CA: Learning Concepts.

Royster, J. D. (1981). Changing employers' attitudes and behaviors toward disabled persons through participative workshops. *Dissertation Abstracts International, 43*(01), 236B. (UMI No. AAD82–09486)

Russ-Eft, D. (1997). Behavioral modeling. In L. J. Bassi & D. Russ-Eft (Eds.), *What works: Training and development practices* (pp. 105–149). Washington, DC: American Society for Training and Development.

*Russ-Eft, D., Krishnamurthi, S., & Ravishankar, L. (1994). Getting results with interpersonal skills training. In J. J. Phillips (Ed.), *In action:*

*Measuring return on investment*. Washington, DC: American Society for Training and Development.

Russ-Eft, D., & Zenger, J. H. (1995). *Behavior modeling: A research summary.* San Jose, CA: Zenger-Miller.

*Russell, J. S. (1982). A study of variables which can increase the effectiveness of behavior modeling training for supervisors and middle managers. *Dissertation Abstracts International, 43*(09), 3096A. (UMI No. AAD83–03847)

*Russell, J. S., Wexley, K. N., & Hunter, J. E. (1984). Questioning the effectiveness of behavior modeling training in an industrial setting. *Personnel Psychology, 40,* 455–588.

Salas, E., & Cannon-Bowers, J. A. (2001). The science of training: A decade of progress. *Annual Review of Psychology, 52,* 471–499.

*Salas, E., Fowlkes, J. E., Stout, R. J., Milanovich, D. M., & Prince, C. (1998). *Does team training improve teamwork skills in the cockpit?* Manuscript submitted for publication.

*Schaab, N. A. (1985). Comparison of two modes of instruction and two modes of behavioral rehearsal in behavior modeling training: An exploration of Graf's generation effect in learning. *Dissertation Abstracts International, 46*(6), 2106B. (UMI No. 8519021)

Schmidt, F. L., & Hunter, J. E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods, 1,* 199–223.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3,* 207–217.

Shute, V. J., & Gawlick, L. A. (1995). Practice effects on skill acquisition, learning outcome, retention, and sensitivity to relearning. *Human Factors, 37,* 781–803.

*Simon, S. J., & Werner, J. M. (1996). Computer training through behavior modeling, self-paced, and instructional approaches: A field experiment. *Journal of Applied Psychology, 81,* 648–659.

*Smith, P. E. (1976). Management modeling training to improve morale and customer satisfaction. *Personnel Psychology, 29,* 351–359.

*Smith-Jentsch, K. A., Salas, E., & Baker, D. P. (1996). Training team performance-related assertiveness. *Personnel Psychology, 49,* 909–936.

*Stacey, J. W. (1987). A comparative study of the effectiveness of a mixed vs. a positive only model in the management/supervisory training technique, behavior modeling. *Dissertation Abstracts International, 49*(03), 409A. (UMI No. AAD88–05971)

*Stearns, C. E. (1984). The effects of self-modeling, behavior rehearsal and traditional modeling upon assertive behavior performance. *Dissertation Abstracts International, 44*(12), 3946B. (UMI No. 8406322)

Stevens, C. K., & Gist, M. E. (1997). Effects of self-efficacy and goal orientation training on negotiation skill maintenance: What are the mechanisms? *Personnel Psychology, 50,* 955–978.

*Stoffey, R. W. (1990). *The effect of behavior modeling training on appraisee and appraiser behavior: Impact on fairness and accuracy of performance appraisal ratings and procedures.* Unpublished doctoral dissertation, Stevens Institute of Technology, Hoboken, NJ.

*Stout, R. J., Salas, E., & Fowlkes, J. E. (1997). Enhancing teamwork in complex environments through team training. *Group Dynamics, 1,* 169–282.

*Stout, R. J., Salas, E., & Kraiger, K. (1997). The role of trainee knowledge structures in aviation team environments. *International Journal of Aviation Psychology, 7,* 235–250.

*Tan, L. C. (1982). *The Filipino behavior model: Adaptations, innovations, and results*. (Available from Development Dimensions International, 1225 Washington Pike, Pittsburgh, PA 15243–0379)

*Taylor, L. K. (1987). An interaction analysis of behavior modeling to facilitate interpersonal supervisory skills of school principals. *Dissertation Abstracts International, 48*(04), 802A. (UMI No. AAD87–10862)

Thornton, G. C. I. (1980). Psychometric properties of self-appraisals of job performance. *Personnel Psychology, 33,* 264–271.

Tulving, E., & Thompson, D. M. (1971). Retrieval processes in recognition memory: Effect of associative context. *Journal of Experimental Psychology, 87,* 116–124.

Tulving, E., & Thompson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80,* 352–373.

Werner, J. M., O'Leary-Kelly, A., Baldwin, T. T., & Wexley, K. N. (1994). Augmenting behavior-modeling training: Testing the effects of pre- and post-training interventions. *Human Resource Development Quarterly, 5,* 169–183.

Wexley, K. N., & Baldwin, T. T. (1986). Posttraining strategies for facilitating positive transfer: An empirical investigation. *Academy of Management Journal, 29,* 503–520.

Wexley, K. N., & Latham, G. P. (2002). *Developing and training human resources in organizations* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Wexley, K. N., & Nemeroff, W. F. (1975). Effectiveness of positive reinforcement and goal setting as methods of management development. *Journal of Applied Psychology, 60,* 446–450.

Wickens, C. D., & Seidler, K. S. (1997). Information access in a dual-task context: Testing a model of optimal strategy selection. *Journal of Experimental Psychology: Applied, 3,* 196–215.

*Woods, C. Q., & Taunton, R. L. (1991). *Effect of targeted intervention on nurse manager leadership*. Unpublished manuscript, University of Kansas Medical Center, Kansas City.

*Yi, M. Y., & Davis, F. D. (2001). Improving computer training effectiveness for decision technologies: Behavior modeling and retention enhancement. *Decision Sciences, 32,* 521–544.

*Zenger-Miller. (1985). *Research summary: Supervision*. (Available from Zenger-Miller, 1735 Technology Drive, San Jose, CA 95110)

*Zorica, J. A. (1988). The design, implementation, and evaluation of a model training program to teach oral communication skills to line managers. *Dissertation Abstracts International, 49*(06), 2405B. (UMI No. AAD88–15223)

Received April 17, 2003
Revision received July 15, 2004
Accepted August 9, 2004 ■