# Factors That Influence
# Skill Decay and Retention:
# A Quantitative Review and Analysis

Winfred Arthur, Jr.

*Department of Psychology*
*Texas A&M University*

Winston Bennett, Jr.

*AL/HRTD, Armstrong Laboratory, Brooks AFB*

Pamela L. Stanush, and Theresa L. McNelly

*Department of Psychology*
*Texas A&M University*

This article presents a review of the skill retention and skill decay literature that focuses on factors that influence the loss of trained skills or knowledge over extended periods of nonuse. Meta-analytic techniques were applied to a total of 189 independent data points extracted from 53 articles. Results indicate that there is substantial skill loss with nonpractice or nonuse, with the amount of skill loss ranging from an effect size (*d*) of –0.01 immediately after training to a *d* of –1.4 after more than 365 days of nonuse. Most of the study's hypotheses for moderators were supported. Physical, natural, and speed-based tasks were less susceptible to skill loss than cognitive, artificial, and accuracy-based tasks. Additionally, certain methodological variables, such as using recognition tests, using similar conditions of retrieval at retention, and using behavioral evaluation criteria, resulted in less skill loss over time. Implications of the results for training and future research are discussed.

The objective of this article is to review the skill decay and skill retention literature in an attempt to delineate the effects of factors that influence the retention of trained

skills over extended periods of nonuse. This was accomplished by using meta-analytic procedures to (a) provide quantitative "population" estimates (i.e., aggregating across multiple primary studies) of the magnitude of skill loss over periods of nonpractice or nonuse; (b) identify major factors that influence skill decay and determine whether, meta-analytically speaking, they are moderators; and (c) attempt to furnish quantitative estimates of the magnitude of the effects of these factors and evaluate their impact on training outcomes. Specifically, meta-analytic procedures were used to rank the identified moderator variables in terms of their influence on skill decay.

*Skill decay* refers to the loss or decay of trained or acquired skills (or knowledge) after periods of nonuse. Skill decay is particularly salient and problematic in situations where individuals receive initial training on knowledge and skills that they may not be required to use or exercise for extended periods of time. Reserve personnel in the military, for example, may be provided formal training only once or twice a year. When called up for active duty, however, it is expected that they will need only a limited amount of refresher training, if any, to reacquire any skill that has been lost and subsequently to perform their mission effectively (Wisher, Sabol, Hillel, & Kern, 1991).

There have been several reviews of the skill decay/retention literature, examples of which include Annett (1979); Farr (1987); Gardlin and Sitterley (1972); Hagman and Rose (1983); Hurlock and Montague (1982); Naylor and Briggs (1961); Prophet (1976), and Schendel, Shields, and Katz (1978). These reviews have all been qualitative in nature, and although they differ in terms of their breadth, depth, comprehensiveness, and sophistication, they are consistent in identifying a core set of major factors that influence the decay or retention of trained skills over extended periods of nonuse. These factors are (a) the length of the retention interval, (b) the degree of overlearning, (c) certain task characteristics (e.g., closed-looped vs. open-looped tasks, physical vs. cognitive tasks), (d) methods of testing for original learning and retention, (e) conditions of retrieval, (f) instructional strategies or training methods, and (g) individual differences. A synthesis and review of these factors are presented later.

In this review, we identify and draw a distinction between methodological and task-related factors. Methodological factors are those that can be modified in the training or learning context to reduce skill loss. Examples of these factors include degree of overlearning, conditions of retrieval, evaluation criteria, and the method of testing. Task-related factors, on the other hand, are inherent characteristics of the task and are typically not amenable to modification by the trainer, researcher, or both. Examples of task-related factors include characteristics such as the distinction between closed-loop and open-looped tasks, physical and cognitive tasks, and natural and artificial tasks. Our results and their implications for future research and practice are discussed within the context of this distinction.

## RETENTION INTERVAL

The effect of progressive knowledge or skill deterioration when knowledge and skills are not used or exercised for extended periods of time is a fairly robust phenomenon. Although the vast majority of the literature consists of laboratory studies (e.g., 88% of the data points analyzed here), in applied settings, this is related to infrequent opportunities to practice or perform acquired skills (Ford, Quiñones, Sego, & Speer Sorra, 1992; Noe, 1986; Peters & O'Connor, 1980). Skill or knowledge loss has also been associated with absent or inadequate feedback (Driskell, Willis, & Copper, 1992; Farr, 1987; Hurlock & Montague, 1982).

Although various skill components may differ in their resistance to decay, in general, the longer the period of nonuse, the greater will be the decay (Annett, 1979; Farr, 1987; Gardlin & Sitterley, 1972; Hurlock & Montague, 1982; Naylor & Briggs, 1961; Prophet, 1976). It is important to note that, although the length of the retention interval (i.e., the nonpractice period) has been cited as a powerful factor in retention (e.g., Annett, 1979; Farr, 1987; Gardlin & Sitterley, 1972; Hurlock & Montague, 1982; Naylor & Briggs, 1961; Prophet, 1976), it is, albeit, a factor that may operate through mechanisms other than time per se (Naylor & Briggs, 1961). Furthermore, it has been argued by some (e.g., Naylor & Briggs, 1961) that the amount of decay is influenced by both task and situational factors.

In this meta-analysis, it was expected that the length of the nonpractice interval would be positively associated with the level of skill decay. Specifically, longer retention intervals were expected to result in more skill decay than shorter retention intervals. It was also expected that this relation would be moderated by factors such as (a) the degree of overlearning, (b) certain task characteristics, (c) the method of testing for original learning and retention, (d) the conditions of retrieval, (e) the instructional strategies or training methods used, (f) speed versus accuracy criteria tasks, and (f) the evaluation criteria used.

## DEGREE OF OVERLEARNING

The single most important determinant of both skill and knowledge retention appears to be the amount or degree of overlearning (Farr, 1987; Hurlock & Montague, 1982; Schendel et al., 1978; Wright, 1973). Overlearning provides additional training beyond that required for initial proficiency. Subsequently, a greater degree of learning is achieved. Several reasons have been proposed to explain the enhancing effect of overlearning on long-term retention. Overlearning may strengthen the bonds between stimulus and response, decreasing the likelihood that the response will decay or be forgotten (Schendel & Hagman, 1982). Additionally, the increased repetitions and practice may provide further feedback to the trainee regarding the correctness of responses and may allow for practice of

performance to improve the correctness of the response. Overlearning probably also enhances automaticity and subsequently reduces the amount of concentrated effort demanded of the trainee. Furthermore, it has been demonstrated that overlearning gives the trainee more confidence in his or her performance and decreases factors (e.g., stress and anxiety) that hamper performance during retention tests (Martens, 1974).

Thus, decay can be reduced or delayed by overlearning. For instance, Schendel and Hagman (1982) found that the degree of initial task overlearning was negatively related to the amount of skill decay on the disassembly and assembly of the M60 machine gun. Results of Driskell et al.'s (1992) meta-analysis of the effects of overlearning on retention also indicated that overlearning produces a significant moderate overall effect on retention. Specifically, Driskell et al. found that the effect of overlearning on retention was moderated by the amount of overlearning, type of task, and the length of the retention period. In this meta-analysis, it was expected that the degree of overlearning would be negatively associated with skill decay such that higher degrees of overlearning would result in less skill decay over periods of nonuse.

## TASK CHARACTERISTICS

Another set of variables that influence the retention and decay of skill and knowledge is the characteristics of the task being learned. A wide range of tasks has been used in the skill decay literature. These tasks differ considerably in terms of difficulty, complexity, and level of integration. Most studies investigating the influence of task characteristics on skill decay have attempted to classify the variety of tasks into broad categories (Farr, 1987). Hence, it seems reasonable to postulate that each of the underlying skill requirements of particular task types could differentially affect the rate of learning and skill retention.

Typically, the skill retention/decay literature includes the following broad classifications of tasks: physical/cognitive and closed-looped/open-looped tasks (Naylor & Briggs, 1961). Other reviews, however, have investigated distinctions of task characteristics that include natural/artificial and integrated/nonintegrated tasks (Annett, 1979) and instrument/contact tasks (Prophet, 1976). Some task characteristics (e.g., task difficulty; see Mumford, Weeks, Harding, & Fleishman, 1987), although recognized as being important to skill decay and retention, have generally not been used as characteristics for classification due to the difficulty in operationally defining them. This study investigated the following task classifications as moderators of skill decay: closed-looped/open-looped, physical/cognitive, natural/artificial, and speed/accuracy.

### Closed-Looped Versus Open-Looped Tasks

Closed-looped tasks, such as preflight checks and other fixed-sequence tasks, usually involve discrete responses that have a definite beginning and end. On the

other hand, open-looped tasks, such as tracking and problem solving, typically involve continuous responses that are repeated and do not have a definite beginning or end. The results of primary empirical studies (e.g., Hufford & Adams, 1961; Mengelkoch, Adams, & Gainer, 1960; Smith & Matheny, 1976) are fairly conclusive in demonstrating that open-looped tasks are better retained, even for extended time periods (months or years), than closed-looped tasks. This consensus is also reflected in the narrative reviews of the literature (e.g., Childs & Spears, 1986; Farr, 1987; Hurlock & Montague, 1982).

The finding that closed-looped tasks generally decay faster than open-looped tasks has been hypothesized to result from the nature of the task. For example, the typical continuous nature of open-looped tasks, which may allow for repeated practice (and thus overlearning) of individual trials, is hypothesized to make such tasks more resistant to decay (Adams, 1967; Naylor & Briggs, 1961). Specifically, individual trials in an open-looped task may be unclear, and thus, repetition of individual trials may occur during performance (Adams, 1967; Naylor & Briggs, 1961). An additional reason for the difference in skill decay between closed-looped and open-looped tasks is that open-looped tasks may be more integrated or coherent than closed-looped tasks and thus may be retained better. Lastly, the two types of tasks may differ in the way skill decay/retention is measured. For example, in the context of motor tasks, it has been suggested that the measurement of closed-looped tasks may be more sensitive to slight performance deviations than those used to measure the retention of open-looped responses (Schendel et al., 1978). In this meta-analysis, it was expected that open-looped tasks would display less skill decay than closed-looped tasks.

## Physical Versus Cognitive Tasks

In their meta-analysis of the effects of overlearning on skill retention, Driskell et al. (1992) drew a distinction between two types of tasks, namely physical and cognitive tasks. *Physical tasks* are characterized as those requiring activities such as muscular strength, exertion of forces, endurance, and coordination. *Cognitive tasks*, on the other hand, involve perceptual input, mental operations, problem solving, and decision making. The conceptual basis for this distinction is based on past research (see Hagman & Rose, 1983; Melnick, 1971) that suggests the type of task may moderate the effect of overlearning on skill retention.

It is argued (e.g., Ryan & Simons, 1981) that cognitive skills should be better retained because they more readily lend themselves to mental practice (i.e., cognitive rehearsal of a task in the absence of overt physical movement), which diminishes the amount of skill loss. In support of this, the results of Driskell, Copper, and Moran's (1994) meta-analysis showed that although mental practice was effective for both cognitive and physical tasks, the effect of mental practice on retention was significantly stronger for cognitive tasks.

Thus, mental rehearsal or imaginary practice appears to be an effective strategy to reduce forgetting or skill decay during the retention interval (Farr, 1987) and is the basis for the thesis that cognitive tasks should suffer less decay than physical tasks. It is, however, important to note that this is based on the premise that there is some mental rehearsal. In fact, contrary to this treatise, it is reasonable to hypothesize that in the absence of mental rehearsal, physical skills should be retained better than cognitive skills.

In this study, studies that involved mental rehearsal during the retention interval were not included because mental rehearsal is considered "practice" of knowledge or skill. Because one of the criteria for inclusion in this meta-analysis was that the retention interval had to be one of nonuse and nonpractice of the skill or knowledge of interest, any study that included mental rehearsal during the retention interval was excluded. It is important to note that very few studies used mental rehearsal as a manipulation. Consequently, for this meta-analysis, it was expected that physical tasks would display less skill decay than cognitive tasks.

## Natural Versus Artificial Tasks

Natural and artificial tasks differ on two major dimensions that might affect retention, namely, complexity and motivation. First, natural tasks are generally more complex. They are, therefore, more elaborately processed, which has a major positive influence on how well they are learned and, subsequently, retained. In other words, the more cohesive or integrated a task is or the more inherently amenable it is to learner-imposed organization (characteristics more common to natural than artificial tasks), the less the skill will decay. Second, in the use of natural tasks, participants generally have a genuine interest in acquiring and retaining proficiency, something that is difficult to ensure with artificial tasks (Annett, 1979). In fact, the role of motivation in skill retention appears to be indirect. Thus, although there is ample recognition that motivation is extremely important to the development of expertise (e.g., Ericsson & Charness, 1994), it is not assumed to play a direct role in retention, but rather, on how much practice people engage in, which, in turn, can be expected to affect retention (via mechanisms such as degree of original learning and organization of material). For instance, Driskell et al. (1992) hypothesized that the effect of overlearning on retention ought to be greater for natural tasks than for artificial tasks, although they were unable to test it in their meta-analysis because they had only one study in their database that used a natural task.

Examples of natural tasks appearing in the literature have included typewriting (Hill, 1957; Swift, 1906; Towne, 1922), simulated lunar landing (Cotterman & Wood, 1967), instrument flying (Mengelkoch et al., 1960), a range of military tasks (McDonald, 1967), and piano playing (Rubin-Rabson, 1939, 1940a, 1940b, 1941a, 1941b, 1941c, 1941d). Examples of artificial tasks have included various forms of tracking (Battig, Nagel, Voss, & Brogden, 1957; Hammerton, 1963; Jahnke, 1958;

Melton, 1964; Trumbo, Noble, Cross, & Ulrich, 1965; Trumbo, Noble, & Swink, 1967), mazes (McGeoch, 1932; McGeoch & Melton, 1929; Tsai, 1924), and a variety of gymnastic skills, such as ball tossing and balancing (Meyers, 1967; Purdy & Lockhart, 1962; Roehrig, 1964; Ryan, 1962, 1965). The results of these studies suggest that natural tasks are generally retained better than artificial tasks. In this meta-analysis, it was expected that natural tasks would display less skill decay than artificial tasks.

## SPEED VERSUS ACCURACY

Speed (e.g., time to complete a task) and accuracy (e.g., number of errors) are two types of criteria that have been used as dependent variables in skill decay studies. The distinction between speed and accuracy as indicators of performance has been compared with the quantity versus quality distinction in the organizational psychology literature (Campbell, 1990). This distinction, however, has not been typically investigated or discussed in the previous reviews of the skill decay literature. One exception is a review of motor tasks (i.e., National Aeronautics and Space Administration space-flight skills), which found that the ability to perform motor tasks in a specified period of time tends to deteriorate more rapidly than performance accuracy (Bodilly, Fernandez, Kimbrough, & Purnell, 1986). Across task types, however, it was expected that speed tasks would display less skill decay than accuracy tasks. This seemingly contradictory prediction is based on the observation that accuracy is considered to be a deficient criterion because learning and skill acquisition have been demonstrated to continue beyond the point of perfect accuracy (Regian & Schneider, 1990). More importantly, accuracy also asymptotes rapidly in many tasks, leading to a potentially false conclusion that the material is mastered when this is in fact not the case.

## METHODS OF TESTING FOR ORIGINAL LEARNING AND RETENTION

The typical paradigm for testing retention usually involves training individuals to some initial criterion on a specified task or skill and testing for performance on the task after some period of nonuse. The test mode for retention can take one of two forms—either using a recall test or a recognition test. The literature indicates that recall and recognition are, in many instances, independent processes such that an individual's ability to recognize an event is unrelated to their ability to recall it (Flexser & Tulving, 1978; Tulving & Weisman, 1975). Hence, different retention measures can yield different degrees of apparent retention, with recall tests usually yielding lower scores than recognition tests (Farr, 1987; Luh, 1922). Consequently, when examining the effect of different variables on skill retention, it is important and essential to recognize the role of the retention test mode or technique as a

potential moderator. An extension of this position is that in comparing the long-term effectiveness of different training protocols, the same retention technique should be used across the different protocols (Farr, 1987). In this meta-analysis, it was expected that studies that used recognition tests would report less skill decay than those that used recall tests.

## CONDITIONS OF RETRIEVAL:
## SIMILARITY OF ORIGINAL LEARNING AND
## RETENTION TEXTING CONTEXTS

Skill retention, in terms of amount and quality, appears to depend on two related factors, namely how information was encoded and the types of cues present at retrieval (Barclay, Bransford, Franks, McCarrell, & Nitsch, 1974; Morris, Bransford, & Franks, 1977). The *encoding specificity principle* (Tulving, 1983; Tulving & Thomson, 1973) states that information retrieval or retention will be maximized if the conditions at retention assessment match as closely as possible to those present during the original learning. Consequently, another factor that has been found to influence retention test scores is the context of testing. Similarity between the condition or context of the recall situation (the retention environment) and those of original learning (the learning environment) allows the stimuli of the learning environment to provide cues that enhance retrieval of information from memory (Cann & Ross, 1989; Godden & Baddeley, 1975; Light & Carter-Sobell, 1970; Schab, 1990; Smith, Glenberg, & Björk, 1978). These rich memory-retrieval cues, along with fewer irrelevant cues, reduce interference during the memory process and improve retrieval of relevant skill and knowledge, thereby decreasing decay or forgetting (Hurlock & Montague, 1982; Naylor & Briggs, 1961). Thus, skill retention scores tend to be higher if the retention measurement is conducted in a context similar to that of the original learning (Driskell et al., 1992; Farr, 1987; Hurlock & Montague, 1982). In applied contexts, the important variables are the functional similarity of the training device (original learning) to the actual job equipment (retention test; Schendel et al., 1978). In terms of long-term retention, the appearance of a training device may be much less important than whether the trainee's performance when using the device is representative of the performance required by the task (Grimsley, 1969a). Furthermore, in summarizing the results of their meta-analysis, Driskell et al. (1992) note that training, within the context of long-term retention, must consider the environmental conditions in which the actual performance will take place. Thus, consistent with the context-dependent memory research, skill retention scores should increase as the similarity between the original learning and retention testing contexts increases (Farr, 1987). Consequently, it was expected that the level of skill decay would be negatively associated with the level of similarity between the original learning and retention testing contexts such that higher levels of similarity would result in lower levels of decay.

## EVALUATION CRITERIA

Evaluation of training programs can be categorized on the basis of four levels or types of criterion measures—reactions, learning, behavior, and results—delineated by Kirkpatrick's (1959, 1987) typology. Reaction criteria measure trainees' feelings or impressions of training. Learning criteria are measures of the learning outcomes of training and are used to assess the knowledge or skill gained by the trainees. Although learning measures can take the form of performance tests and peer evaluations, most learning measures are paper-and-pencil tests measuring the knowledge attained during training (Wexley & Latham, 1991). Behavioral criteria are measures of actual on-the-job performance. On-the-job appraisal can be collected from a variety of sources, including supervisors, coworkers, subordinates, or all of these. Results criteria provide an indication of training-program utility assessed in terms of the contribution of training to organizational objectives such as lower costs, reduced absenteeism and turnover, and company profits. Behavioral and results criteria can be further categorized according to whether they are objective criteria (e.g., number of goods produced) or subjective criteria (e.g., performance ratings by a supervisor).

Although Kirkpatrick's typology is most often used to categorize training evaluation criteria (Tannenbaum & Yukl, 1992), past research and reviews of the skill decay/retention literature have not addressed the relation between skill decay and type of evaluation criterion used. This study attempts to address this limitation in the current literature by examining the influence of evaluation criterion type—specifically learning and behavior criteria—on the amount of skill decay.

Our analysis was limited to learning and behavior criteria because not all criteria types in Kirkpatrick's typology are appropriate in investigations of skill decay/retention. For example, reactions are an affective response to training and are not measures of knowledge, ability, or skill; thus, this criterion type was considered to be inappropriate in an investigation of skill decay. Likewise, results criteria are measures of program utility and contribution to organizational goals and objectives and consequently are not relevant or appropriate in an investigation of skill decay. Hence, neither reaction nor results criteria were included in this meta-analysis. No specific hypotheses were postulated for whether learning or behavior criteria would result in greater or less skill decay over time.

## INSTRUCTIONAL STRATEGIES AND TRAINING METHODS

In any investigation of long-term skill retention, the relation between initial skill acquisition and subsequent retention is vitally important and needs to be taken into account. Specifically, for knowledge and skills to be retained, they must first be acquired via some medium of learning. In a training environment, the instructional

process is the means by which knowledge and skills are acquired. The focus of instructional design and techniques is to facilitate the acquisition of knowledge and skills in the training environment to be transferred later to a second performance environment, typically the job.

Many types of instructional strategies have been used to facilitate the acquisition of skills and knowledge. In training contexts, these instructional strategies are referred to as training methods. Instructors can use a variety of media and techniques to aid in trainees' learning of particular skills or knowledge. The two most frequently used methods of instruction in job-training environments are on-the-job training and the lecture method (Bennett & Arthur, 1997; Goldstein, 1993). Other instructional techniques include programmed instruction, an instructional technique that systematically presents information to the learner. Recent developments in instructional strategies often involve computer-assisted instruction. Audiovisual techniques such as television and films are often used in training environments to facilitate the learning process. Other types of instructional strategies include videodisk technology, machine simulators, team training, and behavior modification.

Although not much attention has been given to the role of instructional strategies and training methods in the skill decay literature, the limited research tentatively suggests that the choice of training method can influence the retention of skills (Ainsworth, 1979). For instance, programmed instruction, usually used for training or intellectual skills, has been found to lead to better retention than conventional (platform-based, lock-step) instruction (Farr, 1987). Although it had been originally planned to include instructional strategies in the meta-analysis, it was not possible to do so because the level of information presented in the primary studies was not specific enough to permit the coding of this variable. That is, there was a paucity of information about the specific instructional strategy or training method used in the acquisition of skills. Nevertheless, this variable is discussed here for the sake of completeness.

## INDIVIDUAL DIFFERENCES

Although there has been some research investigating the role of individual differences in skill decay/retention, it appears that the role of this factor is often confounded by the degree of original learning. Specifically, although results consistently find that higher ability individuals retain more knowledge or skill over periods of nonuse than lower ability individuals, it is argued that higher ability trainees really acquire more knowledge, skills, or both in the same amount of time than lower ability trainees (Farr, 1987; Schendel et al., 1978). Thus, initial skill acquisition is confounded with retention, and the true relation between individual differences and rates of skill decay is difficult to determine. There is some evidence, however, that lower ability learners forget larger portions of abstract, theoretical

material than do higher ability individuals (Farr, 1987). Given the inability to control for the previously mentioned confounding of variables, individual differences as a moderating factor were not included in the meta-analysis. In addition, individual differences have not received a lot of attention in the skill decay literature, so there are few studies investigating this variable (Farr, 1987). This variable, however, is discussed here for the sake of completeness.

## WHY A META-ANALYSIS

A feature common to previous reviews of the skill decay/retention literature (e.g., Annett, 1979; Farr, 1987; Gardlin & Sitterley, 1972; Hagman & Rose, 1983; Hurlock & Montague, 1982; Naylor & Briggs, 1961; Prophet, 1976; Schendel et al., 1978) is that they are all narrative reviews and, subsequently, qualitative in nature. Although narrative reviews are unquestionably meaningful in their own right, there are certain limitations to this method of integrating large literature bases that can be readily addressed by a more quantitative and standardized procedure such as a meta-analysis. In fact, the problems inherent in conducting qualitative reviews are aptly reflected in Schendel et al.'s (1978) comment that

> conflicting data and data pertinent to a more detailed understanding of the behavioral consequences of an extended no-practice period generally were *skimmed over* [italics added] to lend coherence to this report. In doing so, an oversimplified picture of long-term motor memory and the variables that may affect it has been sketched. (p. I)

Extensive discussions of the advantages of quantitative reviews over narrative reviews can be found in such publications as Arthur, Barrett, and Alexander (1991); Glass, McGaw, and Smith (1981); Green and Hall (1984); and Hunter and Schmidt (1990).

In summary, a meta-analytic integration of the skill retention literature is not only possible, it is warranted because it can accomplish objectives that cannot be realized with qualitative reviews including the ability to investigate relations that were not addressed in the primary studies. Thus, the objective of this study was to use meta-analytic procedures to generate aggregate effect sizes across all pertinent studies with the intention of providing a gauge of the overall magnitude of skill loss over periods of nonpractice, nonuse, or both. We also sought to examine the extent to which skill decay is influenced by theoretically relevant and practically important factors.

## METHOD

### Literature Search

An extensive literature search was conducted to identify empirical studies that had investigated skill decay or retention. This process started with a search of nine

computer databases (*Defense Technical Information Center, Econlit, Educational Research Information Center, Government Printing Office, National Technical Information Service, PsychLit, Social Citations Index, Sociofile,* and *Wilson*). The following key words were used: skill acquisition, skill decay, skill degradation, skill deterioration, skill maintenance, skill perishability, skill retention, training effectiveness, training efficiency, and training evaluation. The electronic search was also supplemented with a manual search of the current literature. Approximately 3,600 citations were obtained as a result of this initial search. A review of the abstracts of these citations for appropriate content (i.e., empirical studies that actually investigated skill decay or retention), along with a decision to retain only English language articles, narrowed the list down to 172 articles. In addition, the reference lists of these articles were reviewed, and a number of researchers in the area were contacted to try to obtain additional published and unpublished studies. As a result of these efforts, an additional 98 articles were identified, resulting in a total of 270 articles. Each article was then reviewed and considered for inclusion in the meta-analysis. The sources of these articles were as follows: journal articles (48%), technical reports (41%), books/book chapters (4%), conference papers and presentations (4%), dissertations (1%), masters theses (1%), and unpublished or submitted manuscripts (1%).

## Inclusion Criteria

A number of decision rules were used to determine the data points (studies) that would be included or retained for the meta-analysis. First, to be included in the meta-analysis, a study must have investigated skill loss or retention over time with an identifiable interval of nonuse or nonpractice between the acquisition and retention test session. Thus, the studies had to report both preretention and postretention performance data. Second, tasks or skills were limited to "organizationally related" tasks or complex skill acquisition. Thus, for example, training interventions that focused on parent training (e.g., Therrien, 1979) were excluded. Furthermore, studies that used children as participants (e.g., Kittel, 1957; Shuell & Keppel, 1970) were also excluded. Third, to be included in the meta-analysis, a study had to report sample sizes along with an outcome statistic (e.g., univariate $F$, $t$, $\chi^2$,) or other pertinent information (e.g., group means and standard deviations) that allowed the computation of or conversion to a $d$ statistic using the appropriate conversion formulas (see Glass et al., 1981; Hunter & Schmidt, 1990; Wolf, 1986).

Using these decision rules resulted in a retention of 52 (19%) of the 270 articles. The reasons for excluding some studies were as follows: no retention interval or not a nonpractice/nonuse retention interval (32%), insufficient statistical information to calculate or convert results to $d$ (32%), nonempirical or not a primary study (26%), use of children/nonadult participants (4%), nonorganizational study (e.g., parent training; 3%), and unable to locate or obtain a copy of the article (3%).

## Data Set

*Nonindependence.*    As a result of the inclusion criteria, an initial data set of 249 data points (*ds*) from 52 articles was obtained. Some of the data points, however, were nonindependent. Effect sizes or data points are nonindependent if they are computed from data collected on a single group of participants. Decisions about nonindependence have to also take into account whether the effect sizes represent the same variable or construct or not.

For a number of reasons, nonindependence is an important consideration when conducting a meta-analysis. First, one effect of nonindependence is to reduce the observed variability of the effect sizes. Under these conditions, interpretations of the homogeneity of effect sizes must be made very cautiously. Another effect of nonindependence is to artificially inflate sample sizes and effects beyond the number of independent data points. Although this may increase the power of the meta-analysis, it becomes difficult to determine the amount of error in the statistics describing the data points. A final effect of nonindependence is to overweight the contribution (either positively or negatively) of the studies or articles contributing multiple nonindependent data points. Consequently, to address these problems, when data points are nonindependent, the accepted practice is to aggregate them by finding the average. Implementing this practice resulted in 189 independent data points from 52 articles.

*Outliers.*    A number of prominent statisticians have noted that virtually all data sets contain at least some outlier data points (Gulliksen, 1986; Mosteller & Hoaglin, 1991; Tukey, 1960, 1977). Because meta-analyses sometimes include "studies of imperfect methodological quality, the presence of outliers is highly probable" (Schmidt et al., 1993, p. 10). Thus, an outlier in the meta-analytic framework would be a primary study effect size that does not appear to be consistent with the other study effect sizes, either because of errors in the data collection or computation or because of some very unusual feature of the study design or choice of participants. Detecting outliers in meta-analytic data sets is potentially very important because the effect of such outliers is typically an increase in the residual variability and a possible shift in the mean effect size.

Huffcutt and Arthur's (1995) sample-adjusted meta-analytic deviancy (SAMD) statistic was computed for each data point to detect outliers. In Huffcutt and Arthur's procedure, outliers or extreme data points are identified using a scree plot (Dillon & Goldstein, 1984; Loehlin, 1987) to set a cutoff above which data points are considered to be outliers. Specifically, the absolute values of the SAMD statistics are rank ordered from the highest to the lowest and plotted. SAMD values that rise above the flat gradual slopes are identified as potential outliers and are investigated.

SAMD statistics were computed across all 189 *d*s. The mean SAMD value was −0.19 (*SD* = 3.84). The resulting SAMD scree plot is presented in Figure 1. As this chart indicates, the first 11 data points appear to rise above the flat portion of the plot and thus were identified as outliers. The absolute SAMD values of data points identified as outliers ranged from 7.80 to 19.32. A follow-up analysis and a detailed review suggested that the deviancy could be attributed to unusual study features in several of the outlier studies. The 11 outliers constituted 5.82% of the 189 *d*s in the data set. Dropping the 11 outliers resulted in a final data set of 178 independent *d*s. The sources of these data points were as follows: journal articles (75%), technical reports (20%), dissertations (4%), and unpublished or submitted manuscripts (1%). The references for these sources are listed in the reference section and are preceded by an asterisk.
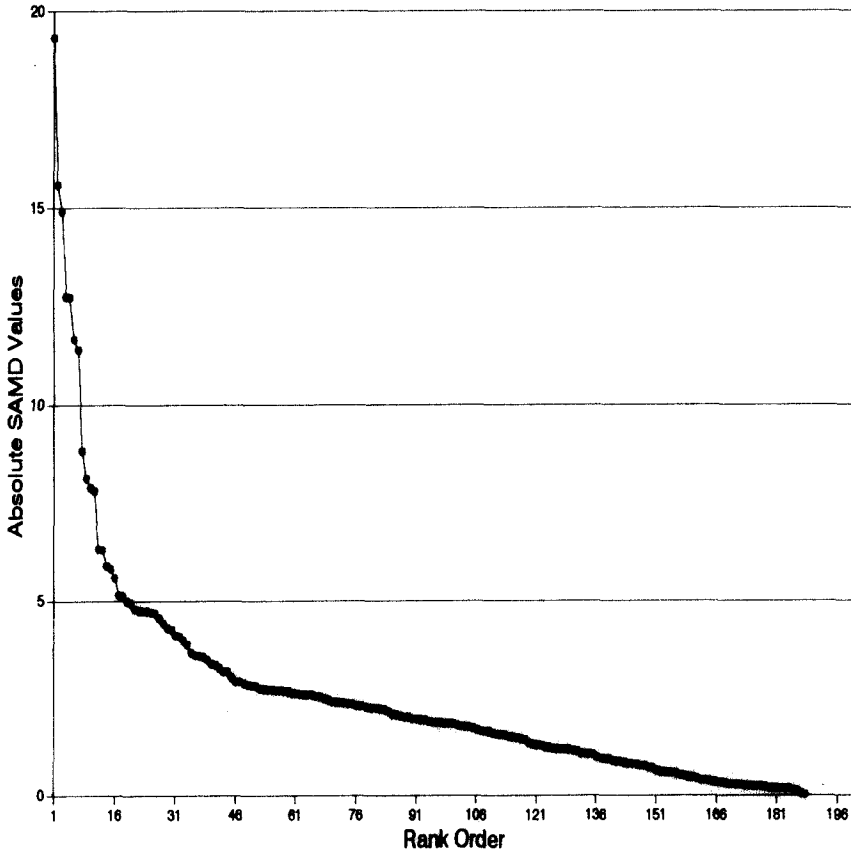


FIGURE 1    Scree plot for sample-adjusted meta-analytic deviancy statistics.

## Description of Variables

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

*Retention interval.*    This was coded as the number of days between the end of original training or learning and the test for retention. As such, this variable represented the length of the nonpractice interval, nonuse interval, or both. There were some articles that reported a retention interval range (e.g., "the retention test was administered 60–75 days after the original training session"). For these articles, the retention interval was coded as the midpoint of the range (e.g., 67.5 days).

*Degree of overlearning.*    Overlearning refers to the deliberate overtraining of a task past a set criterion performance level. In the typical overlearning paradigm, a task criterion may be set at one errorless trial. Participants in the control condition practice the task until performance reaches the criterion level. Participants in the treatment condition practice the task until they reach this level and then receive additional practice trials. For example, if reaching the criterion level takes 10 trials, the overlearning manipulation may constitute an additional 5 trials (50% overlearning), an additional 10 trials (100% overlearning), or other degrees of overlearning. Retention is then assessed at some interval after the training session. Driskell et al.'s (1992) operationalization of degree of overlearning (DOV) was used in the this study where

$$DOV = \frac{\% \text{ learning in higher condition}}{\% \text{ learning in higher condition}\% + \text{learning in lower condition}}$$

Thus, a DOV value of zero indicates there was no overlearning.

*Task characteristics.*    Using the definitions presented in earlier sections of this article, tasks were coded on the following dimensions: (a) closed-looped versus open-looped, (b) physical versus cognitive, (c) natural versus artificial, and (d) speed versus accuracy criteria.

*Method of testing for original learning and retention.*    Both the original learning and retention tests were coded as being either recall or recognition tests.

*Conditions of retrieval—similarity of original learning and retention testing contexts.*    The similarity between the retention measurement and original learning was coded as a dichotomous variable (i.e., either similar or different). To be coded as similar, the two contexts (i.e., the original learning and retention testing)

had to be the same. That is, participants had to be brought back to the same location, and the protocol from the first testing session had to be the same as that used in the original testing session. If the retention testing session was changed in some way from that used in the original testing session (e.g., retesting in a different location), the testing sessions were coded as "different."

*Evaluation criteria.* Two of Kirkpatrick's (1959, 1987) evaluation criteria types were coded. Specifically, both original learning and retention criteria were coded as being either learning or behavioral in nature.

## Coding Accuracy and Interrater Agreement

Pamela L. Stanush and Theresa L. McNelly coded the data reported in this meta-analysis. The coding training process and implementation were as follows: First, the coders were furnished with a copy of a coder training manual and a reference guide, which had been developed by Winfred Arthur, Jr. and Winston Bennett, Jr. and used with other meta-analysis projects. Each coder used the manual and reference guide to code a single article on their own. Next, they attended a follow-up training meeting with all the authors to discuss problems encountered in using the guide and the coding sheet and to make changes to the guide, the coding sheet, or both as required. They were then assigned the same five articles to code. After coding these five articles, the coders attended a second training session in which the degree of convergence between them was assessed. Discrepancies and disagreements related to the coding of the five articles were resolved using a consensus discussion and agreement among all four authors.

After this second meeting, the articles used in the meta-analysis were individually assigned to the coders for coding. As part of this process, the coders coded a common set of 20 articles that were used to assess the degree of interrater agreement. Interrater agreement was assessed by comparing the values recorded by each coder for each of the variables of interest. Raters were in agreement if identical values were recorded by both coders. The level of agreement obtained for the primary meta-analysis variables is presented in Table 1. As these results indicate, the level of agreement was generally high, with a mean overall agreement of 96.67% ($SD = 3.12$).

## Calculating the Effect Size Statistic

In meta-analysis, cumulating the effects across studies requires that outcomes from all studies be converted to a common metric (Hunter & Schmidt, 1990). This study

TABLE 1
Interrater Agreement for Major Study Variables

| Variable | Agreement |
| --- | --- |
| d | 95% |
| N | 90% |
| Retention interval | 95% |
| Degree of overlearning | 95% |
| Task characteristics | |
|    Closed-looped vs. open-looped | 95% |
|    Physical vs. cognitive | 100% |
|    Natural vs. artificial | 100% |
|    Speed vs. accuracy | 95% |
| Methods of testing | 100% |
| Conditions of retrieval (similarity) | 95% |
| Evaluation criteria | |
|    Learning criteria | 100% |
|    Behavior criteria | 100% |
| Overall | 96.67% |

used the effect size statistic ($d$) as the common metric. The effect size, or $d$ statistic, provides a measure of the strength of a treatment or independent variable (e.g., different training methods). The effect size statistic, $d$, is the standardized difference between two means. Thus, in experimental designs, it represents the observed difference between the experimental and the control group in standard-deviation units (Cohen, 1990). A positive $d$ value indicates that the experimental group performed better than the control group on the dependent variable. Conversely, a negative $d$ value indicates that the control group performed better than the experimental group, and a zero $d$ value indicates no difference between the groups. Cohen (1992) described small, medium, and large effect sizes as 0.20, 0.50, and 0.80, respectively. Thus, a medium effect size represents half a standardized difference between means.

As shown in Equation 1, the $d$ statistic is calculated as the difference between the means of the experimental ($M_E$) and control groups ($M_C$) divided by a measure of the variation (Cohen, 1988; Glass, 1976; Glass et al., 1981; Hunter & Schmidt, 1990).

$$d = \frac{M_E - M_C}{S_W} \tag{1}$$

The measure of variation used in this study, $S_W$, is the pooled, within-group standard deviation (Hunter & Schmidt, 1990).

Although Equation 1 calls for the means of "experimental" and "control" groups, the typical skill retention or decay primary study did not use control groups (or even pretraining data). The most frequently used paradigm was to train a group to some criterion, test the group, and collect performance data immediately after training, and then test again after a specified interval of nonuse. The difference in perform- ance between the two testing occasions (original learning and retention) thus represented the amount of skill loss.

In calculating the $d$s in this meta-analysis, the original learning performance (i.e., amount of skill acquired immediately after training and before the retention interval) was used as the control group score ($M_C$), and performance on the retention test was used as the experimental group score ($M_E$). Due to skill deterioration, the amount of skill remaining after a retention interval is typically smaller than the skill attained immediately following training (i.e., before the retention interval). As a result, the effect sizes calculated are usually negative; that is, the original learning score immediately following training (control-group performance) was better than the retention score (experimental-group performance). A negative $d$ in this meta- analysis, then, indicates that skill has deteriorated during the retention interval; the larger the negative $d$, the more the skill has decayed. A positive $d$ indicates that the amount of skill attained immediately following training increased during the retention interval (an unusual finding unless practice or rehearsal occurred during the retention period). A $d$ value of zero indicates that there was no loss of skill during the retention interval.

For studies that reported actual means and standard deviations for retention and original learning performance (77%), effect sizes were calculated directly using these statistics. For studies that reported other statistics (e.g., correlations, $t$ statis- tics, or univariate two-group $F$ statistics; 23%), the appropriate conversion formulas (Dunlap, Cortina, Vaslow, & Burke, 1996; Glass et al., 1981; Hunter & Schmidt, 1990; Wolf, 1986) were used to convert them to $d$s.

## Analyses

*Cumulating effect sizes across studies.* Using Arthur, Bennett, and Huffcutt's (1995; Huffcutt, Arthur, & Bennett, 1993) SAS PROC MEANS meta- analysis program, mean sample-weighted effect sizes ($\bar{d}$) were calculated using Equation 2 following:

$$\bar{d} = \frac{\sum d_i \, n_i}{N_T} \qquad (2)$$

where $d$ is the mean effect size; $d_i$ is the effect size for each study; $n_i$ is the sample size for each study; and $N_T$ is the total sample size across all studies. Sample weighting assigns studies with larger sample sizes more weight and reduces the effect of sampling error because sampling error generally decreases as the sample size increases (Hunter & Schmidt, 1990).

As previously indicated, the $d$ statistic is a standard deviation metric used to express the difference between treatment and control groups, usually in experimental studies. There may be instances where the sample sizes are very uneven. In the context of the typical skill retention paradigm, this may be due to attrition during the nonpractice/nonuse retention interval. In these situations, Hunter and Schmidt (1990) recommend "correcting" the mean d ($\bar{d}$) for the attenuating effect of unequal or unbalanced sample sizes. This is accomplished using a bias multiplier, denoted as "A," which is calculated as (Hunter & Schmidt, 1990, pp. 281–283, 289):

$$A = 1 + (.75 / \bar{N} - 3)) \tag{3}$$

where $\bar{N}$ is the average sample size across studies. It should be noted that for sample sizes of 100 or larger, the bias multiplier will differ only trivially from 1.00. The corrected mean $d$ ($\bar{\delta}$) and standard deviation of the population effect sizes (SD$\delta$) are then obtained by dividing the mean $d$ and standard deviation by the bias multiplier as presented in Equation 4 and Equation 5:

$$\delta = \bar{d} / A \tag{4}$$

$$SD\delta = Var(\delta)^{\frac{1}{2}} / A \tag{5}$$

where $Var(\delta)$ is the population variance.

*Moderator analyses.*    For the assessment of each factor proposed to influence skill retention, studies were categorized into separate subsets according to the specified level of the factor. An overall, as well as a subset mean effect size, was then calculated for each factor. A moderator variable or factor was identified if (a) the effect size variance was lower in the subsets than the factor as a whole, (b) the average effect size varied from subset to subset, or (c) both preceding conditions were present. In brief, if large differences were found between subsets of a given factor, then the factor could be considered to be a moderator variable.

# RESULTS

## Retention Interval

The first research objective was to assess the effect of the length of the nonpractice interval on the amount of skill decay. It was hypothesized that the length of the nonpractice interval would be positively associated with the amount of skill decay such that longer retention intervals would result in more skill decay than shorter retention intervals. Because several (45%) of the intervals had single data points, the coded intervals were categorized into eight groups before analyzing the data. The categorization scheme used was rational in nature and was intended to reflect an exponential increase in retention intervals. The eight time intervals and the number of days they represent are presented in Table 2. The correlation between these eight retention intervals and the original time interval was .73 ($p = .00005$).

The results of the meta-analysis to test the first research objective, which are presented in Table 3, indicate that there is an increase in the amount of skill decay as the length of the nonpractice interval increases. Furthermore, although it was based on only eight data points, the correlation between retention interval and corrected mean $d$ ($\delta$) was –0.51. Consistent with the hypothesis, studies with longer retention intervals reported more skill loss. The standard deviations of the corrected mean $d$s ($SD\delta$) reported in Table 3, however, are large enough to suggest the presence and operation of potential moderator variables. Thus, as also hypothesized, it would seem that the nature of the skill-loss/nonpractice interval relation is influenced by moderating factors. To test for these moderators, separate meta-analyses were run for the subsets of these variables.

The results of the moderator analysis, which are presented in Table 4, indicate that most of the factors may be operating as moderators (i.e., the mean $d$s vary from subset to subset, and the variances are lower in the subsets). It is important, however, to note that these analyses are collapsed across all retention intervals. Thus, these

TABLE 2
Retention Interval Categories

| Retention Intervals | Number of Days |
| --- | --- |
| 1 | Less than 1 day |
| 2 | Greater than or equal to 1 day; less than or equal to 7 days |
| 3 | Greater than 7 days; less than or equal to 14 days |
| 4 | Greater than 14 days; less than or equal to 28 days |
| 5 | Greater than 28 days; less than or equal to 90 days |
| 6 | Greater than 90 days; less than or equal to 180 days |
| 7 | Greater than 180 days; less than or equal to 365 days |
| 8 | Greater than 365 days |

## TABLE 3
### Results of Overall Meta-Analysis for the Relationship Between Retention Interval And Skill Retention

| Retention Interval | Number of Data Points | Total Sample Size | Corrected statistics | | | Variance Due to Sampling Error | Min d | Max d | 95% Confidence Interval | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | | Mean d | δ | SDδ | | | | Lower Bound | Upper Bound |
| Overall | 178 | 8719 | -0.97 | -0.95 | 0.72 | 15.39 | -5.24 | 0.99 | -2.35 | 0.44 |
| 1 | 3 | 45 | 0.02 | 0.02 | 0.00 | 100.00 | -0.12 | 0.17 | 0.02 | 0.02 |
| 2 | 89 | 3325 | -1.03 | -1.01 | 0.77 | 17.02 | -5.24 | 0.82 | -2.53 | 0.51 |
| 3 | 13 | 680 | -1.54 | -1.52 | 0.63 | 20.20 | -2.87 | -0.02 | -2.75 | -0.28 |
| 4 | 19 | 766 | -0.95 | -0.93 | 1.00 | 10.04 | -4.06 | 0.59 | -2.90 | 1.03 |
| 5 | 33 | 2656 | -0.65 | -0.65 | 0.39 | 25.35 | -1.84 | 0.99 | -1.42 | 0.12 |
| 6 | 7 | 302 | -1.42 | -1.39 | 0.82 | 14.83 | -3.33 | -0.22 | -3.00 | 0.22 |
| 7 | 11 | 670 | -1.04 | -1.03 | 0.39 | 32.89 | -2.67 | -0.55 | -1.80 | -0.26 |
| 8 | 3 | 275 | -1.28 | -1.27 | 0.35 | 30.19 | -2.48 | -0.92 | -1.95 | -0.58 |

## TABLE 4
### Meta-Analysis Results for Moderator Analyses (Collapsed Across all Retention Intervals)

| Moderator Variable | Number of Data Points | Total of Sample Size | Corrected Statistics | | | | | | 95% Confidence Interval | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Mean d | δ | SDδ | % Variance Due to Sampling Error | Min d | Max d | Lower Bound | Upper Bound |
| Overall | 178 | 8719 | -0.97 | -0.95 | 0.72 | 15.39 | -5.24 | 0.99 | -2.35 | 0.44 |
| Degree of overlearning | | | | | | | | | | |
| 60% | 10 | 520 | -1.34 | -1.32 | 0.60 | 21.11 | -2.49 | -0.18 | -2.48 | -0.15 |
| 66.7% | 10 | 376 | -1.55 | -1.52 | 0.63 | 26.32 | -2.62 | 0.82 | -2.74 | -0.29 |
| 67% | 7 | 356 | -1.37 | -1.35 | 1.12 | 7.28 | -4.98 | -0.16 | -3.54 | 0.84 |
| 75% | 2 | 36 | -1.30 | -1.24 | 0.00 | 100.00 | -1.31 | -1.29 | -1.24 | -1.24 |
| 78.9%[a] | 1 | 9 | -1.32 | — | — | — | — | — | — | — |
| Task characteristics | | | | | | | | | | |
| Closed–Looped | 58 | 2372 | -0.73 | -0.71 | 0.61 | 22.22 | -4.06 | 0.99 | -1.65 | 0.16 |
| Open–Looped | 120 | 6347 | -1.06 | -1.04 | 0.74 | 14.13 | -5.24 | 0.82 | -2.47 | 0.38 |
| Physical | 69 | 4399 | -0.76 | -0.75 | 0.48 | 22.75 | -2.87 | 0.72 | -1.69 | 0.19 |
| Cognitive | 109 | 4320 | -1.18 | -1.15 | 0.83 | 14.73 | -5.24 | 0.99 | -2.98 | 0.48 |
| Natural | 67 | 4748 | -0.94 | -0.93 | 0.68 | 12.02 | -4.98 | 0.72 | -2.26 | 0.41 |
| Artificial | 111 | 3971 | -1.00 | -0.98 | 0.75 | 18.69 | -5.24 | 0.99 | -2.44 | 0.48 |
| Speed | 33 | 637 | -0.33 | -0.32 | 0.44 | 52.55 | -2.37 | 0.99 | -1.18 | 0.55 |
| Accuracy | 145 | 8082 | -1.02 | -1.00 | 0.71 | 14.08 | -5.24 | 0.82 | -2.39 | 0.38 |

(Continued)

Table 4
(Continued)

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Methods of testing | | | | | | | | | | |
|   Recognition | 19 | 1160 | -0.86 | -0.85 | 0.24 | 55.34 | -2.15 | 0.00 | -1.33 | -0.38 |
|   Recall | 159 | 7559 | -0.98 | -0.96 | 0.76 | 14.26 | -5.24 | 0.99 | -2.45 | 0.52 |
| Conditions of retrieval | | | | | | | | | | |
|   Similar | 174 | 8621 | -0.95 | -0.94 | 0.70 | 15.53 | -5.24 | 0.99 | -2.31 | 0.44 |
|   Different | 4 | 98 | -2.14 | -2.07 | 0.47 | 54.02 | -3.33 | -1.57 | -2.99 | -1.14 |
| Evaluation criteria | | | | | | | | | | |
|   Learning | 157 | 5624 | -1.07 | -1.04 | 0.82 | 16.09 | -5.24 | 0.99 | -2.65 | 0.57 |
|   Behavioral | 21 | 3095 | -0.78 | -0.77 | 0.36 | 18.63 | -2.37 | 0.37 | -1.48 | -0.07 |

[a]Meta-analysis results are not presented for this degree of overlearning because a single data point cannot be meta-analyzed.

analyses do not take into account potential interactions between the moderators and the retention interval. The reasons for the absence of fully hierarchical moderator analyses and the effects of this are noted in the Discussion.

## Degree of Overlearning

It was hypothesized that the DOV would be negatively associated with skill decay such that higher degrees of overlearning would result in less skill decay over periods of nonuse. Because 83% ($N = 148$) of the data points did not have an overlearning manipulation or report any information on the degree of overlearning, the test of this hypothesis was limited to only those data points that reported information on the degree of overlearning. The results presented in Table 4 indicate that there was a fairly limited range in the degree of overlearning used in the studies in the data set. Consistent with this, the differences in the amount of skill decay as a function of degree of overlearning were also limited.

## Task Characteristics

*Closed-looped versus open-looped tasks.*    It was hypothesized that open-looped tasks would display less skill decay than closed-looped tasks. The results of the moderator analysis presented in Table 4 indicate that the closed-looped/open-looped task distinction is operating as a moderator. The δs for these moderators are different from the overall (−0.95) with a higher overall level of retention being reported for closed-looped (−0.71) instead of open-looped tasks (−1.04)—a finding that is in contrast to the study hypothesis.

*Physical versus cognitive tasks.*    It was hypothesized that physical tasks would display less skill decay than cognitive tasks. The results of the moderator analysis presented in Table 4 indicate that across all retention intervals the amount of skill decay for physical tasks ($\delta = -0.75$) is less than that for cognitive tasks ($\delta = -1.15$). Thus, the study hypothesis was supported—physical tasks display less skill decay than cognitive tasks, and the difference in decay is close to half a standardized unit (i.e., 0.40) across all retention intervals.

*Natural versus artificial tasks.*    It was hypothesized that natural tasks would display less skill decay than artificial tasks. The results of the moderator analysis presented in Table 4 indicate that across all retention intervals, the amount of skill

decay for natural tasks ($\delta$ = −0.93 is only slightly less than that for artificial tasks ($\delta$ = −0.98). Thus, although the difference in magnitude was small, the results suggest that natural tasks are less susceptible to decay than artificial tasks. The study hypothesis was, therefore, supported.

*Speed versus accuracy tasks.*   It was hypothesized that speed tasks would display less decay than accuracy tasks. The results of the moderator analysis presented in Table 4 indicate that across all retention intervals, the amount of skill decay for accuracy tasks was over three times higher than that of speed tasks (i.e., $\delta$ = −1.00 and −0.32, respectively).

## Methods of Testing for Original Learning and Retention

It was hypothesized that studies that used recognition tests would report less skill decay than those that used recall tests. Because there were no studies in the data set that switched from one type of test to the other from the original learning to retention test, the results presented in Table 4 are based on studies that used the same type of test (i.e., either recognition or recall) for both the original and retention test. As shown in Table 4, although the difference is fairly small (0.11), the use of recognition tests resulted in less decay ($\delta$ = −0.85) than recall tests ($\delta$ = −0.96); this is consistent with the study hypothesis.

## Conditions of Retrieval—Similarity of Original Learning and Retention Test Context

It was hypothesized that the level of skill decay would be negatively associated with the level of similarity between the original learning and retention contexts such that higher levels of similarity would result in less skill decay. In support of this hypothesis, the results presented in Table 4 indicate a large difference (1.13) between the two conditions. It should be noted, however, that the results for different context are based on only four data points, which is a rather small number from a meta-analytic perspective.

## Evaluation Criteria

This factor was included in this meta-analysis to investigate whether the outcomes of skill retention studies are influenced by the type of criterion used to measure

both original learning and retention performance. The type of evaluation criterion used, namely learning and behavior, was coded for both original learning and for retention criteria. The type-of-criterion data for original learning and retention were identical, however. In other words, all studies that used learning criteria for the assessment of original learning also used learning criteria for the assessment of retention. The same was true for the use of behavioral criterion measures. Subsequently, the results presented later are for both original learning and retention criteria.

The results of the moderator analysis presented in Table 4 indicate that the evaluation criterion type moderates skill retention. The δs and variances vary as a function of criteria type and also differ from the overall effects. Generally speaking, the level of skill decay was less for behavioral criteria ($\delta = -0.77$) compared with learning criteria ($\delta = -1.04$). Although a hypothesis was not postulated for this variable, this finding could be explained by the nature of the task that each type of criteria is intended to measure. Behavioral measures are usually measures of on-the-job performance after training. Hence, they are more likely to be natural tasks performed in applied settings than artificial tasks performed in laboratory settings. And, because natural and applied tasks have been demonstrated to be less susceptible to skill loss, one would subsequently also expect behavioral criteria to manifest higher levels of retention. This is in contrast to learning criteria, which are more likely to be posttraining "classroom"-type tests and, thus, are more likely to be laboratory-based tasks performed in artificial settings. Table 5 presents a frequency breakdown of criterion type by the natural/artificial distinction and clearly demonstrates this to be the case.

## Relative Influence of Moderators on Skill Decay

The final objective of the this study was to rank order the identified moderators in terms of their relative influence on skill decay. The meta-analytically generated effect sizes were compared with the judgmental ratings of effect reported by Farr (1987). Table 6 presents a rank order (descending) of all the moderators investigated in this study along with the ratings assigned by Farr (1987). The δs presented for the meta-analysis moderators represent the absolute difference between the δs for the levels of that specified moderator. Farr (1987) reviewed a number of narrative

TABLE 5
Frequency Break Down of Evaluation Criteria by the Natural/Artificial Distinction

| Characteristic | Learning Criteria | Behavioral Criteria |
| --- | --- | --- |
| Natural | 32% | 76% |
| Artificial | 68% | 24% |

**TABLE 6**

Comparison of Relative Importance of Moderators (Collapsed Across all Retention Intervals) Based on the Present Meta-Analysis and Farr's (1987) Ratings Based on Other Narrative Reviews

| Moderator Variable | Absolute $\delta$ Difference Between Levels | I | II | III | IV | V | VI | VII | Mean[a] |
|---|---|---|---|---|---|---|---|---|---|
| Conditions of retrieval | 1.13 | — | 3 | — | 4 | 5 | — | — | 4 |
| Speed/accuracy | 0.68 | — | — | — | — | — | — | — | — |
| Physical/cognitive | 0.40 | — | — | — | — | — | — | — | 3.8 |
| Closed—/open—looped | 0.33 | 2 | — | — | 5 | 4 | 4 | 4 | 3.8 |
| Evaluation criteria | 0.27 | — | 2 | — | 5 | — | — | — | — |
| Retention interval[b] | 0.20 | 4 | — | — | — | 5 | 4 | 4 | 3.8 |
| Methods of testing | 0.11 | — | 3 | 4 | 4 | 4 | 4 | 5 | 5 |
| Overlearning[c] | 0.08 | 4 | — | — | — | — | 4 | — | 4 |
| Natural/artificial | 0.05 | — | — | — | — | — | — | — | — |

*Note.* I = Annett (1979); II = Gardlin & Sitterly (1972); III=Hagman & Rose (1983); IV=Hurlock & Montague (1982); V=Naylor & Briggs (1961); VI=Prophet (1976); VII=Schendel et al. (1978).

[a]Mean of Farr's (1987) ratings. Ratings are on a 0 (*weakest effect*) to 5 (*strongest effect*) scale. [b]This represents the correlation between length of the retention interval and the amount of skill loss that has been converted to a *d*. [c]This represents the correlation between degree of overlearning and the amount of skill loss that has been converted to a *d*; this data is limited to only those studies that reported information on the degree of overlearning ($k = 30$).

reviews and, on the basis of this, assigned a rating to each factor. The ratings were intended to represent the "effect" of the variable "on the course of forgetting or skill decay" (p. 46). The strength of the effect is Farr's judgment of how "clearly, strongly, and consistently the variable affects retention in a predictable way" (p. 46). The "strongest effect" was rated a 5, and the "weakest effect" was assigned a zero (see Farr, 1987, Tables A2–A7, pp. A6–A28).

The data presented in Table 6 indicate that not all the meta-analysis moderators investigated in this study were included in Farr's (1987) ratings. Several variables were discussed in previous narrative reviews but were not rated by Farr (e.g., speed/accuracy and natural/artificial), whereas other variables included in this meta-analysis have not been addressed in any previous reviews (e.g., the evaluation criterion-type classification).

For those variables that are common to both this meta-analysis and Farr's (1987) ratings, the rankings reveal little convergence between the two. The variable with the largest absolute difference in the meta-analysis was conditions of retrieval, whereas the highest ranking variable in Farr's (1987) rating (i.e., 5) was methods of testing. The lowest ranking variable for the meta-analysis was the natural/artificial distinction, which was not included in Farr's (1987) ranking. On the other hand, Farr's (1987) lowest rankings were for closed-looped/open-looped distinction and retention interval.

Two other fairly obvious discrepancies are for methods of testing and overlearning, which Farr (1987) rated as having large effects but, for the meta-analysis, resulted in small absolute differences. For methods of testing, Farr reported that "the particular retention measure used can affect the degree of retention found" (pp. A–22). The methods of retention compared in the meta-analysis were recognition versus recall, so it is unclear whether Farr's comparison was referring to a broader class of retention measures compared with the comparison made by this meta-analysis. This could help explain the discrepancy between the two. As mentioned earlier, the small number of data points collected for overlearning in the meta-analysis made interpretation of this variable somewhat problematic.

In conclusion, the results of the meta-analysis generally provide quantitative support for the majority of trends noted by Farr's (1987) qualitative review of the skill decay literature. The major point of departure between the two, however, concerns the relative ranking of the major variables influencing skill decay/retention. Farr's ranking of the effect of each variable on a rating scale ranging from 0–5 did not (closely) match the relative ranking of obtained effect sizes obtained by this meta-analysis.

## DISCUSSION

Skill loss during periods of nonuse is particularly problematic in situations where individuals receive initial training on knowledge and skills that they may not be

required to use or exercise for extended periods of time, as exemplified by military reservists and other personnel (e.g., disaster teams). Consistent with past research, the results of the meta-analysis indicate that the relation between skill retention and the length of the nonpractice or nonuse interval is negative. The meta-analysis, however, went one step further by providing a quantitative population estimate of the magnitude of this relation aggregated across multiple extant primary studies. Specifically, it demonstrated that the amount of skill loss ranges from a $d$ of $-0.1$ immediately after training (less than one day) to a $d$ of $-1.4$ after more than 365 days of nonuse. That is, after more than 365 days of nonuse or nonpractice, the average participant was performing at less than 92% of their performance level before the nonpractice interval.

The results of this study also demonstrated that several important factors moderate the skill decay/nonpractice interval relation. These included both variables that have been discussed in past primary studies and review articles (e.g., degree of overlearning; closed-looped/open-looped, physical/cognitive, natural/artificial, and speed/accuracy tasks; methods of testing; and conditions of retrieval) and others that were specifically identified in this study (e.g., evaluation criteria). Quantitative estimates of the magnitude of the effects of these factors along with their impact on training outcomes were also demonstrated.

Most of the study's hypotheses for the moderators were supported. Specifically, for task-related factors, physical, natural, and speed-based tasks were less susceptible to skill loss than cognitive, artificial, and accuracy-based tasks. On the other hand, for the methodological factors, using recognition tests, similar conditions of retrieval at retention, and behavioral evaluation criteria resulted in less skill loss than using recall tests, different conditions of retrieval at retention, and learning evaluation criteria.

It was hypothesized that open-looped tasks would display less skill decay than closed-looped tasks. This was the only hypothesis that was not supported because closed-looped tasks displayed less decay than open-looped tasks. This finding is inconsistent with past results of primary empirical studies (e.g., Hufford & Adams, 1961; Mengelkoch et al. 1960; Smith & Matheny, 1976) and narrative reviews of the skill decay literature (e.g., Childs & Spears, 1986; Farr, 1987; Hurlock & Montague, 1982) that have demonstrated that open-looped tasks are better retained than closed-looped tasks over extended periods of nonuse. One plausible explanation for the inconsistent findings may be the presence of an "interaction" between moderator variables that may be contaminating the results. An examination of the distribution of data points across closed-looped/open-looped tasks and physical/cognitive tasks reveals an unbalanced distribution of data points across the levels of these two moderator variables (see Table 7). Although physical tasks are represented fairly equally across open-looped versus closed-looped tasks (34 and 35 data points, respectively), this is not the case for cognitive tasks, which are represented by 24 data

TABLE 7
Breakdown of Number of Data Points by Moderator Variable

| | Closed-Looped | | | | Open-Looped | | | |
|---|---|---|---|---|---|---|---|---|
| | Natural | | Artificial | | Natural | | Artificial | |
| | Speed | Accuracy | Speed | Accuracy | Speed | Accuracy | Speed | Accuracy |
| Physical | 15 | 17 | 1 | 1 | 0 | 9 | 3 | 23 |
| Cognitive | 0 | 11 | 5 | 8 | 0 | 15 | 9 | 61 |

points for closed-looped tasks but 85 data points for open-looped tasks. Because the (absolute) observed mean $d$ effect size is substantially larger for cognitive tasks ($d = -1.18$) than for physical tasks ($d = -0.76$), the mean $d$ obtained for open-looped tasks may be artificially inflated due to the overrepresentation of cognitive tasks. A similar imbalance is also present on the speed–accuracy distinction. Again, whereas speed and accuracy tasks are relatively equally distributed across closed-looped tasks (21 and 38 points, respectively), this is not the case for open-looped tasks (12 and 108 points, respectively). And because the (absolute) observed mean $d$ is substantially larger for accuracy tasks ($d = -1.02$) compared with speed task ($d = -0.33$), as in the physical/cognitive distinction, this may again explain why open-looped tasks appear to be displaying more skill loss over time than closed-looped tasks.

The use of meta-analytic procedures also allowed an empirical assessment of the relative effect of the identified moderators on skill decay. This would appear to be an improvement over past attempts to rate judgmentally the effect of moderators as they relate to the phenomenon of skill decay (e.g., Farr, 1987). The results of this study indicate that the similarity of the conditions of retrieval was the most important moderator. This result is interesting because conditions-of-retrieval is directly related to the issue of transfer of training. *Transfer of training* is the generalization of trained performance, in a given task, from the training environment to the work environment and is one of the key criteria for evaluating the effectiveness of any formal training program (Kirkpatrick, 1987). In the context of skill decay/retention, however, a nonpractice interval exists between performance in the training and work environments. The results of this study suggest that the similarity of the training (acquisition) and work (retention) environments plays a major role in the retention of skills and knowledge over periods of nonuse or nonpractice, providing additional support for a basic tenant in training-program design—that is, to enhance retention, trainers should try to ensure the functional similarity of both the training device (acquisition) and actual job equipment (retention) and the environment in which both are performed.

The second most important moderator (see Table 6) appeared to be the use of speed versus accuracy-dependent variable tasks, and the least important moderator was the natural/artificial task distinction. The degree of overlearning also seemed to have a relatively weak effect, but this may be misleading for a variety of reasons. First, very few studies (only 30 studies, 17%) reported any information on the degree of overlearning. Second, there was a fairly limited range in the degree of overlearning used in the studies in the data set. This limited range may have attenuated the effect of this variable in the meta-analysis. Third, like the length of the retention interval, the comparative δ is based on a converted correlation instead of the difference between levels. For this reason, the results presented for the length of the retention interval and the degree of overlearning in Table 6 may not be truly comparable to that reported for the other moderators.

This study also sought to assess the relation between skill decay and evaluation criterion type (i.e., learning and behavior criteria). The results indicated that the amount of skill decay was lower for behavioral than for learning criteria.

In summary, this article demonstrates that not only is a meta-analysis of the skill decay/retention literature possible, but it can also be very informative. For instance, the distinction between methodological and task-related factors becomes important when it is demonstrated that such factors influence the susceptibility of learned skills to decay, which this study has shown. Because methodological variables can be modified, researchers and practitioners could focus on and select those methodological variables that appear to increase the likelihood that skill is retained over time. For example, this study found that if the conditions of retrieval were similar to the conditions of skill acquisition, the amount of skill lost is markedly less than when the conditions of retrieval are different from the conditions of skill acquisition.

Obviously, there are some methodological variables that experimenters and practitioners have more control of than others. The criteria used to evaluate the retention of skill, for example, may theoretically be modifiable but, in practice, may be inconvenient or impossible to accomplish. Although changing the training and retention conditions to maximize skill retention over time may not always be feasible, as just mentioned, it is important for researchers and practitioners to consider carefully all the factors related to skill decay and retention before the design and development of training programs and evaluations. Maximizing the potential for skill retention over time saves money and time, both in the applied and research world.

Although task characteristics are variables that are intrinsic to the task being trained and thus are not modifiable, the findings presented here can be used, for instance, as aids in determining and scheduling the frequency and amount of overtraining and refresher training for specified task types (e.g., cognitive tasks) that decay faster than others (e.g., physical tasks) to avoid detrimental losses of skill or knowledge.

## LIMITATIONS

No research study is without its limitations. First, our initial intention was to generate skill retention curves for the results presented in Table 3 and Table 4 in an attempt to test the negatively accelerated function suggested by prior reviews. Testing various fit functions would have allowed us to determine empirically and to provide a population estimate of the skill retention curve. These analyses would also have allowed us to take the retention interval into account in the analyses of the moderators. We were unable, however, to do this because these analyses required the use of parametric statistical procedures (specifically, regressions and correlations) with their associated assumptions of normality. This was a problem because the nature of our meta-analysis data set was such that both $d$s and retention intervals were very nonnormal (interval skewness = 4.24, kurtosis = 20.35; $d$ skewness = $-1.12$, kurtosis = 1.70). Although we could have transformed the data to make it more normal, we considered this to be inappropriate because the meta-analysis data as given represent the state of these variables as they exist in the extant literature. There are also no theories to suggest that their distribution should be normal. Therefore, any transformations to achieve normality would have been an inappropriate distortion of reality. Nevertheless, the distribution of data points across retention intervals was even enough to make the interpretation of the moderator analyses meaningful.

Second, in almost all the moderator analyses, the relatively small amount of variance accounted for, coupled with the size of the standard deviation of δ, suggested the presence of additional moderators. Because the choice of moderators in this study was, however, all theoretically or conceptually driven, the decision was made to not seek out additional moderators on a post hoc basis.

Third, although we considered investigating interactions among the moderators (i.e., fully hierarchical moderator analysis), this was not possible because this calls for dividing the data points into a (cells) matrix based on the number of moderators. The feasibility of fully hierarchical moderator analyses is primarily a function of exactly how many data points there are in each cell, because when the number of data points in each cell are as small as was the case in this study (see Table 7), stability and interpretability of the meta-analytic estimates become a major and serious concern.

Fourth, along these lines, there were several factors and potential moderators that were not included in this study primarily because they are either relatively minor or the pertinent information was impossible to extract from primary studies and subsequently code. The relatively large standard deviations of δ, however, might warrant their inclusion in future meta-analytic research. These variables include (a) miscellaneous task characteristics such as task integration, level of task organization, task structure and complexity, and task difficulty (Annett, 1979; Gardlin & Sitterly, 1972; Hurlock & Montague, 1982; Mumford et al., 1987; Naylor

& Briggs, 1961; Schendel et al., 1978); (b) training characteristics such as distribution of practice (e.g., part vs. whole, massed vs. distributed), programmed learning, memory aids, spacing of trials or sessions, feedback, and hypnosis during training (Annett, 1979; Hurlock & Montague, 1982; Naylor & Briggs, 1961; Schendel et al., 1978); (c) retention interval and test characteristics such as rehearsal, test trial characteristics, kinds of rehearsal, relearning, practice during rehearsal, test taking during retention interval, and repetition of test trials (Annett, 1979; Hurlock & Montague, 1982; Naylor & Briggs, 1961; Schendel et al., 1978); (d) individual differences such as motivation of trainee, amount of previous training, intelligence of trainee, ability of trainee, and trainee age (Annett, 1979; Hurlock & Montague, 1982; Naylor & Briggs, 1961; Schendel et al., 1978); and (e) perceptual skills that involve "the ability to discriminate between and to classify stimuli based on perceivable properties" (Proctor & Dutta, 1995, p. 33). Although it is recognized that the tasks included in this meta-analysis have a perceptual component along with a cognitive and motor component, the acquisition of perceptual skills was not a primary focus of this meta-analysis. This study focused on tasks that might be of interest in organizational settings. Thus, the criteria for inclusion of studies limited tasks to those that were organizationally related or involved complex skill acquisition. Of future interest might be a quantitative compilation of tasks associated primarily with perceptual skill acquisition and decay and an investigation of whether these types of skills have different retention rates compared with those of physical and cognitive tasks. This investigation, however, was beyond the scope of this study.

## IMPLICATIONS AND SUGGESTIONS FOR FUTURE RESEARCH

Several issues were made salient by the meta-analysis, and it is suggested that these issues be addressed, or at least considered, in future investigations of the skill decay phenomena. Briefly, these were issues pertaining to (a) the design of protocols and paradigms to enhance skill retention, (b) the lack of attention given to skill acquisition, (c) the lack of consensus concerning criteria for the end of acquisition and the beginning of the retention interval, (d) the failure to assess level of previous skill or knowledge, (e) the role of motivation and individual differences in skill retention, (f) skill decay in the context of team tasks and skills, and (g) the complete reporting of data in primary studies. Each of these factors is discussed in more detail later.

First, the results of the meta-analysis support the argument made by Naylor and Briggs (1961) that the magnitude of skill loss is specific to the task and situation. This finding has several implications for training programs and the personnel who develop them. For instance, the finding that certain task characteristics are more susceptible to decay indicates that these types of skills may need to be retrained

more frequently if a period of nonuse is expected. Although the task characteristics are usually chosen to represent later performance on the job and thus are not easily interchangeable just because some may be more susceptible to skill decay than others, there are some variables that do seem to have a significant effect on the degree of long-term retention and can be manipulated. In fact, the two variables that this study identified as having the largest effects are not related to training content but instead are related to measurement and methodology (i.e., speed/accuracy distinction and conditions of retrieval). Furthermore, the findings for conditions of retrieval (i.e., whether preretention and postretention conditions were different or similar) suggest that training conditions should be as similar to the retention conditions as possible for maximum retention.

Second, although the skill decay literature is represented by an investigation of a variety of tasks, a range of time intervals, and many research paradigms, there are, unfortunately, several methodological and conceptual shortcomings that make drawing conclusions about the phenomena difficult. One of the most pervasive weaknesses of the skill decay literature is the lack of attention given to the phenomena of skill acquisition. Schmidt and Björk (1992), for example, criticized the educational and training settings for treating learning (i.e., skill acquisition) and retention as two separate phenomena that have been studied independently by different scientists, using different methods in different laboratories. These authors argued that the two are really inseparable and need to be considered together when conducting studies on skill decay. In any investigation of long-term skill retention, the relation between skill acquisition and skill retention is vitally important and needs to be taken into account. For example, a researcher or a training specialist can use all the "best" methods to facilitate retention by manipulating aspects of the retention interval and the retention testing situation, but if little or no skill or knowledge is initially acquired during training, retention as a phenomenon becomes a moot issue. It should be recognized that the quality and quantity of skill acquisition is a significant factor in any investigation of skill decay and long-term retention. Recognizing the problem, however, does not mean that remedying it is easy. Attempting to quantify qualitative aspects of a phenomena is difficult, and skill acquisition is no exception.

Third, another problem with the skill decay literature is the lack of consensus concerning the criteria used to determine the point at which skill acquisition should cease and the retention interval should begin. Many primary studies, for example, have trained individuals to one error-free trial (e.g., Hagman, 1980a, 1980b; Schendel & Hagman, 1982), whereas other studies have used criteria such as a predetermined percentage of students correctly performing the task (e.g., Holgrem, Hilligoss, Swezey, & Enkins, 1979; Shields, Goldberg, & Dressel, 1979) as the point to end skill acquisition. Lastly, some studies did not specify a particular criterion that participants had to reach before skill acquisition was terminated; instead, participants were required to complete a certain amount of training material

(e.g., Adams & Hufford, 1962) or to practice a certain task for a specified amount of time (e.g., Arthur, Day, Bennett, McNelly, & Jordan, 1997).

In addition to different types of criteria used to determine the termination of skill acquisition, differences in terminology are also a problem in the skill decay literature. One errorless trial, for example, has been labeled differently across studies (e.g., "proficiency," Hagman, 1980a; "minimal mastery," Farr, 1987; and "mastery," Hall, Ford, Whitten, & Plyant, 1983). The term *mastery* has also been used to refer to one errorless trial (e.g., Hall et al., 1983), to two error-free trials (e.g., Schendel & Hagman, 1982), and to three error-free trials (e.g., Goldberg, Drillings, & Dressel, 1981).

By establishing a criterion such as one or three errorless trials that individuals must achieve before beginning the retention interval, researchers have attempted to standardize the amount of skill acquired by each individual. There are several problems with this methodology, however. Performance can be measured on several dimensions, of which accuracy is only one criterion. Further, accuracy has been criticized for being a deficient criterion because learning and skill acquisition continue beyond the point that accuracy is perfect, and, more importantly, accuracy asymptotes rapidly in many tasks, leading to a potentially false conclusion that the material has been mastered (Regian & Schneider, 1990). Partial support for this effect was obtained by this meta-analysis, which demonstrated that accuracy tasks are more susceptible to decay than speed tasks. So, although measures of speed change more continuously than accuracy, regardless of the criterion used, there is still the problem of identifying the appropriate cutoff or criterion at which one should end skill acquisition.

Investigations of overlearning are methodologically weak if an inappropriate criterion, such as a certain number of trials or percentage of time past one error-free trial, is arbitrarily defined as mastery or proficiency. Overlearning, in some cases, then, may simply be representing an increased amount of skill acquisition. Because learning continues past error-free trials, overlearning may really just be a higher level of skill acquisition. Hence, in the absence of a clear and standardized operational definition of mastery or proficiency that cuts across all, or at least most studies, these terms may simply be serving as arbitrary, meaningless, and convenient cutoff points at which the measure of overlearning can begin.

Fourth, another problem with the methodology currently used in the extant skill decay research is that there is usually no assessment of previous skill or knowledge. In other words, individuals are usually not tested before training to assess how much relevant material they already know. Bahrick (1979), for·example, found that previous training experience facilitated long-term skill retention even though all individuals exhibited the same level of skill proficiency immediately after training. Bahrick (1979) concluded that criterion performance at the end of training is not a sufficient predictor of long-term retention. It makes intuitive sense that any previous knowledge or skill that has been retained from previous acquisition will be further

reinforced by the current experimental manipulation and will be less likely to be lost compared with knowledge or skills that are being learned for the first time.

Two methods of measuring skill acquisition have also been used in the extant literature—namely how much is trained in a specified amount of time and how long it takes to train a certain amount of material. Although these criteria measure certain dimensions of performance, it cannot be assumed that they are interchangeable.

Fifth, another issue is the role of motivation and individual differences in skill retention. The finding that studies utilizing artificial tasks resulted in more skill loss over time than studies that used natural tasks would suggest that motivation may play a role in determining how much skill is retained over time. It is well documented that motivation plays an important role in learning and performance (Kanfer, 1992). Hence, it seems reasonable to suggest that the motivation to learn may also influence the long-term retention of acquired skills. For instance, motivation might play a role in how much practice people engage in, which in turn can be expected to affect retention via mechanisms such as degree of original learning and organization of material. Additionally, it is recognized that "complex skill acquisition requires sustained task attention and practice—effort that is affected by an individual's interest in the task" (Kanfer, 1992, p. 95). The use of more "real-world" tasks in the study of complex skill acquisition and retention should be seriously considered by future research. Relatedly, motivation should be investigated as a factor that might influence the relation between complex skill acquisition and long-term retention.

Future research should also seriously consider the study of individual differences within the context of skill retention. Although it has generally been argued and demonstrated that higher ability individuals (compared with lower ability individuals) retain more knowledge and skill over periods of nonuse because they acquire more in the same amount of time (Carron, 1971; Carron, & Marteniuk, 1970; Farr, 1987; Fox, Taylor, & Caylor, 1969; Grimsley, 1969b; Purdy, & Lockhart, 1962; Schendel et al., 1978; Vineberg, 1975), there is dissenting research that suggests there is also a qualitative difference between higher and lower ability individuals. This difference may explain the enhanced skill retention exhibited by higher ability individuals. Farr (1987), for example, suggested that the differential decay rates observed between higher and lower ability individuals might be due to higher ability individuals using more effective strategies to acquire knowledge and skills. This is consistent with the findings of Hall et al. (1983), who required Navy sailors to complete two self-paced courses in basic electricity and electronics to a criterion of mastery. After a nonpractice retention interval ranging from 18 to 34 days, Hall et al. found that higher ability sailors retained significantly more than lower ability sailors.

Regardless of one's position, the study of individual differences within the context of skill decay and retention is particularly interesting because individual differences may be useful not only for predicting speed of skill acquisition in

original learning, but also in predicting the rate of skill decay and reacquisition (Christal, 1976). If so, a variety of individual difference predictors and data could be used to identify those less likely to benefit from retraining or less likely to perform effectively after retraining. Individual difference data could also be used to schedule the frequency and length of time between retraining sessions.

Sixth, skill decay in individual-versus-team tasks is another issue worthy of future research. With the recent surge in the use of work teams in organizations (Driskell & Salas, 1992), there has been a concurrent interest in how to train teams effectively to work together (Salas, Bowers, & Cannon-Bowers, 1995; Tannenbaum & Yukl, 1992). Several studies (e.g., Bohlander & McCarthy, 1996; Salas, Dickinson, Converse, & Tannenbaum, 1992) have started to furnish information on the differences between effective and ineffective teams, and others, such as Swezey and Salas (1992), have begun to develop guidelines for team training. There appears, however, to be a complete lack of attention to skill decay in team tasks, and no studies could be identified for this meta-analysis. Evaluating the effectiveness of training protocols in the context of skill loss is a logical extension of any research program or paradigm that seeks to assess the comparative effectiveness of specified training protocols (Arthur et al., 1997; Schmidt & Björk, 1992). As noted by Schmidt and Björk, acquisition and retention are really inseparable and need to be considered together in investigations of skill acquisition. Thus, like individual training, future research should investigate factors that influence skill decay in team training tasks.

And finally, as with many other meta-analyses, it must be noted that the reporting of data in the skill decay literature is poor. Although an initial collection of skill decay studies revealed many empirical investigations, the majority of studies could not be coded due to an insufficient amount of information reported. As a concluding comment, we reiterate that a conscientious effort must be made to report all pertinent information and data in future primary studies. Such information should include, but not be limited to, the pertinent test statistic (e.g., $r$, $t$, or $F$), sample sizes, means, and standard deviations. This information will facilitate the inclusion of more studies in future meta-analyses and will permit the investigation of additional potential moderator variables that we were unable to assess here.

## ACKNOWLEDGMENTS

An earlier version of this article was presented at the 102nd Annual Conference of the American Psychological Association, New York, August 1995.

## REFERENCES

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

Adams, J. A. (1967). *Human memory*. New York: McGraw-Hill.

*Adams, J. A., & Hufford, L. E. (1962). Contribution of a part-task trainer to the learning and relearning of a time-shared flight maneuver. *Human Factors, 4*, 159–170.

Ainsworth, J. S. (1979). *Symbol learning in Navy technical training: An evaluation of strategies and mnemonics* (TAEG Rep. No. 66 AD–AO68 041). Orlando, FL: Training Analysis and Evaluation Group.

Annett, J. (1979). Memory for skill. In M. M. Gruneberg & P. E. Morris (Eds.), *Applied problems in memory*. London: Academic.

Arthur, W., Jr., Bennett, W., Jr., & Huffcutt, A. I. (1995). *A SAS PROC MEANS-based method for conducting meta-analysis research: An operational guide and user's manual*. Unpublished manuscript, Texas A&M University, College Station.

Arthur, W., Jr., Barrett, G. V., & Alexander, R. A. (1991). Prediction of vehicular accident involvement: A meta-analysis. *Human Performance, 4*, 89–105.

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

Bahrick, H. P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. *Journal of Experimental Psychology: General, 108*, 296–308.

*Baldwin, R. D., Cliborn, R. E., & Foskett, R. J. (1976). *The acquisition and retention of visual aircraft recognition skills* (Report No. 76–A–4). Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences.

Barclay, J. R., Bransford, J. D., Franks, J. J., McCarrell, N. S., & Nitsch, K. (1974). Comprehension and semantic flexibility. *Journal of Verbal Learning and Verbal Behavior, 13*, 471–481.

*Battig, W. F., Nagel, E. H., Voss, J. F., & Brogden, W. F. (1957). Transfer and retention of bi-dimensional compensatory tracking after extended practice. *American Journal of Psychology, 70*, 75–80.

Bennett, W., Jr., & Arthur, W., Jr. (1997). *Factors that influence the effectiveness of training in organizations: A review and meta-analysis* (AL–TP–1997–0026). Brooks Air Force Base, TX: AL/HRTD, Technical Training Research Division.

*Bilodeau, E. A., & Blick, K. A. (1965). Courses of misrecall over long-term retention intervals as related to the strength of pre-experimental habits of word association. *Psychological Reports, 16*, 1173–1192.

Bodilly, S., Fernandez, J., Kimbrough, J., & Purnell, S. (1986). *Individual ready reserve skill retention and refresher training options* (Report No. AD–A183 416). Santa Monica, CA: Rand Corporation.

Bohlander, G. W., & McCarthy, K. (1996). How to get the most from team training. *National Productivity Review, 15*, 25–35.

*Boker, R. J. (1974). Immediate and delayed retention effects of interspersing questions in written instructional passages. *Journal of Educational Psychology, 66*, 96–98.

*Braden, S. R. (1924). An extensive experiment in motor learning and relearning. *Journal of Educational Psychology, 15*, 313–315.

*Britton, B. K., & Glynn, S. M. (1987). *Revising instructional texts to improve retention: Tests of signaling, retrieval, readability, and the revisers' hypothesis*. Unpublished manuscript. University of Georgia, Athens.

Campbell, J. P. (1990). Modeling the performance prediction problem in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 1, pp. 687–732). Palo Alto, CA: Consulting Psychologists Press.

Cann, A., & Ross, D. A. (1989). Olfactory stimuli as context cues in human memory. *American Journal of Psychology, 102,* 91–102.

*Carron, A. V. (1971). Effect of ability upon retention of a balance skill after two years. *Perceptual and Motor Skills, 33,* 527–529.

Carron, A. V., & Marteniuk, R. C. (1970). Retention as a balance skill as a function of initial ability level. *Research Quarterly, 41,* 478–483.

*Charness, N., & Campbell, J. E. (1988). Acquiring skill at mental calculation in adulthood: A task decomposition. *Journal of Experimental Psychology: General, 117,* 115–129.

Childs, J. M., & Spears, W. (1986). Flight-skill decay and recurrent training. *Perceptual and Motor Skills, 62,* 235–242.

Christal, R. E. (1976). *What is the value of aptitude tests?* Paper presented at the annual conference of the Military Testing Association, Gulf Shores, AL.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist, 45,* 1304–1312.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159.

Cotterman, T. E., & Wood, M. E. (1967). *Retention of simulated lunar landing mission skills: A test of pilot reliability* (Tech. Rep. No. AMRL–TR–66–222). Dayton, OH: Wright-Patterson Air Force Base.

Dillon, W. R., & Goldstein, M. (1984). *Multivariate analysis.* New York: Wiley.

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

Driskell, J. E., & Salas, E. (1992). Can you study real teams in contrived settings? The value of small group research to understanding teams. In R. W. Swezey & E. Salas (Eds.), *Teams: Their training and performance* (pp. 101–124). Norwood, NJ: Ablex.

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

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1,* 1–8.

Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist, 49,* 725–747.

Farr, M. J. (1987). *The long-term retention of knowledge and skills: A cognitive and instructional perspective.* New York: Springer-Verlag.

*Fisk, A. D., Rogers, W. A., Lee, M. D., Hodge, K. A., & Whaley, C. J. (1991). *Automatic information processing and high-performance skills: Principles of consistency, part-task training, content, retention, and complex task performance* (Interim Tech. Rep. No. AFH RL–TR–90–84). San Antonio, TX: Brooks Air Force Base.

Flexser, A. J., & Tulving, E. (1978). Retrieval independence in recognition and recall. *Psychological Review, 85,* 153–171.

Ford, J. K., Quiñones, M., Sego, D., & Speer Sorra, J. S. (1992). Factors affecting the opportunity to perform trained tasks on the job. *Personnel Psychology, 45,* 511–527.

Fox, W. L., Taylor, J. E., & Caylor, J. S. (1969). *Aptitude level and the acquisition of skills and knowledges in a variety of military training tasks* (HumRRO Tech. Rep. 69–6). Alexandria, VA: Human Resources Research Organization.

Gardlin, G. R., & Sitterley, T. E. (1972). *Degradation of learned skills: A review and annotated bibliography* (D180–15081–1, NASA–CR–128611). Seattle, WA: Boeing.

*Gentile, J. R., Monaco, N., Iheozor-Egiofor, I. E., Ndu, N. A., & Ogbonaya, P. K. (1982). Retention of "fast" and "slow" learners. *Intelligence, 6,* 125–138.

*Gilbert, T. (1957). Overlearning and the retention of meaningful prose. *Journal of General Psychology, 56,* 281–289.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5,* 3–8.

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

Godden, D., & Baddeley, A. D. (1975). Context-dependent memory in two natural experiments: On land and under water. *British Journal of Psychology, 66,* 325–331.

Goldberg, S. L., Drillings, M., & Dressel, J. D. (1981). *Mastery training: Effect on skill retention* (Tech. Rep. No. 513). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Goldstein, I. L. (1993). *Training in organizations.* Pacific Grove, CA: Brooks/Cole.

Green, B. F., & Hall, J. A. (1984). Quantitative methods for literature reviews. *Annual Review of Psychology, 35,* 37–53.

Grimsley, D. L. (1969a). *Acquisition, retention and retraining: Effects of high and low fidelity in training devices* (HumRRO Tech. Rep. No. 69–1). Alexandria, VA: Human Resources Research Office.

*Grimsley, D. L. (1969b). *Acquisition, retention, and retraining: Training category IV personnel with low fidelity devices* (HumRRO Tech. Rep. 69–12). Alexandria, VA: Human Resources Research Organization.

*Gulgoz, S., Britton, B. K., & Glynn, S. M. (1987). *Retention of texts revised by Robert Graves: Tests of Graves' readability, signaling, and retrieval hypothesis.* Unpublished manuscript, University of Georgia, Athens.

Gulliksen, H. (1986). The increasing importance of mathematics in psychological research (Pt. 3). *The Score, 9,* 1–5.

*Hagman, J. D. (1980a). *Effects of training schedule and equipment variety on retention and transfer of maintenance skill* (Research Rep. No. 1309). Alexandria, VA: Army Research Institute.

*Hagman, J. D. (1980b). *Effects of training task repetition in retention and transfer of maintenance skill* (Research Rep. No. ADA1201672XSP). Alexandria, VA: Army Research Institute for Behavioral and Social Sciences.

Hagman, J. D., & Rose, A. M. (1983). Retention of military tasks: A review. *Human Factors, 25,* 199–213.

Hall, E. R., Ford, L. H., Whitten, T. C., & Plyant, L. R. (1983). *Knowledge retention among graduates of basic electricity and electronics schools* (HumRRO Tech. Rep. No. 149). Orlando, FL: Training Analysis and Evaluation Group, Department of the Navy.

*Hammerton, M. (1963). Retention of learning in a difficult tracking task. *Journal of Experimental Psychology, 66,* 108–110.

*Healy, A. F., Fendrich, D. W., & Proctor, J. D. (1990). Acquisition and retention of a letter-detection skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 270–281.

Hill, L. B. (1957). A second quarter century of delayed recall, or relearning at eighty. *Journal of Educational Psychology, 48,* 65–69.

Holgrem, J. R., Hilligoss, R., Swezey, R. W., & Enkins, R. (1979). *Training effectiveness and retention of training extension course instruction in the combat arms* (Research Rep. No. 1208). Alexandria, VA: Army Research Institute.

Huffcutt, A. I., & Arthur, W., Jr. (1995). Development of a new outlier statistic for meta-analytic data. *Journal of Applied Psychology, 80,* 327–334.

Huffcutt, A. I., Arthur, W., Jr., & Bennett, W., Jr. (1993). Conducting meta-analysis using the PROC MEANS procedure in SAS. *Educational and Psychological Measurement, 53,* 119–131.

Hufford, L. E., & Adams, J. A. (1961). *The contribution of part-task training to the relearning of a flight maneuver* (Report No. NAVTRADEVCEN TR-297-2). Port Washington, NY: U.S. Naval Training Device Center.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings.* Newbury Park, CA: Sage.

Hurlock, R. E., & Montague, W. E. (1982). *Skill retention and its implications for navy tasks: An analytical review* (NPRDC Special Rep. No. 82–21). San Diego, CA: Navy Personnel Research and Development Center.

Jahnke, J. C. (1958). Retention in motor learning as a function of amount of practice and rest. *Journal of Experimental Psychology, 55,* 270–273.

*Jones, M. B. (1989). Individual differences in skill retention. *American Journal of Psychology, 102,* 183–196.

Kanfer, R. (1992). Motivation theory and industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 1, pp. 75–155). Palo Alto, CA: Consulting Psychologists Press.

Kirkpatrick, D. L. (1959). Techniques for evaluating training programs. *Journal of the American Society for Training and Development, 13,* 3–9.

Kirkpatrick, D. L. (1987). Evaluation of training. In R. L. Craig (Ed.), *Training and development handbook: A guide to human resource development* (pp. 301–319). New York: McGraw-Hill.

Kittel, J. E. (1957). An experimental study of the effect of external direction during learning on transfer and retention principles. *Journal of Educational Psychology, 48,* 391–405.

*Knerr, C. M., Harris, J. H., O'Brien, D. K., Sticha, P. J., & Goldberg, S. L. (1984). *Armor procedural skills: Learning and retention* (Interim Tech. Rep. No. ADA1532274XSP). Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences.

*Konoske, P. J., Ellis, J. A. (1991). *Cognitive factors in learning and retention of procedural tasks* (Report No. ADA254898). San Diego, CA: Navy Personnel Research and Development Center.

*Kruegar, W. C. (1929). The effect of overlearning on retention. *Journal of Experimental Psychology, 12,* 71–78.

*Kruegar, W. C. (1930). Further studies in overlearning. *Journal of Experimental Psychology, 13,* 152–155.

*Leavitt, H. J., & Schlosberg, H. (1944). The retention of verbal and motor skills. *Journal of Experimental Psychology, 34,* 404–417.

*Lewis, D., McAllister, D. E., & Adams, J. A. (1951). Facilitation and interference in performance on the modified Mashburn apparatus: I. The effects of varying the amount of original learning. *Journal of Experimental Psychology, 41,* 247–260.

Light, L. L., & Carter-Sobell, L. (1970). Effects of changed semantic context on recognition memory. *Journal of Verbal Learning and Verbal Behavior, 9,* 1–11.

Loehlin, J. C. (1987). *Latent variable models: An introduction to factor, path, and structural analysis.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Luh, C.W. (1922). The conditions of retention. *Psychological Monographs, 31*(3, Whole No. 142).

Martens, R. (1974). Arousal and motor performance. In J. H. Wilmore (Ed.), *Exercise and Sport Sciences Review* (Vol 2, pp. 155–188). New York: Academic.

*McCullagh, P., & Little, W. S (1989). A comparison of modalities in modeling. *Human Performance, 2,* 101–111.

*McDonald, R. D. (1967). *Retention of military skill acquired in basic combat training* (HUMRRO Tech. Rep. 67–13). Alexandria, VA: Human Resource Research Organization.

McGeoch, J. A. (1932). The comparative retention values of a maze habit, of nonsense syllables, and of rational learning. *Journal of Experimental Psychology, 15,* 662–680.

McGeoch, J. A., & Melton, A. W. (1929). The comparative retention values of maze habits and nonsense syllables. *Journal of Experimental Psychology, 12,* 392–414.

*Melnick, M. (1971). Effects of overlearning on the retention of a gross motor skill. *Research Quarterly, 42,* 60–69.

Melton, A. W. (1964). *Retention of tracking skill: Final Report* (ORA Project 02855). Ann Arbor: University of Michigan.

Mengelkoch, R. F., Adams, J. A., & Gainer, C. A. (1960). *The forgetting of instrument flying skills as a function of the level of initial proficiency* (Report No. NAVTRADEVCEN TR–71–16–18). Port Washington, NY: U.S. Naval Training Device Center.

Meyers, J. L. (1967). Retention value of balance learning as influenced by extended layoffs. *Research Quarterly, 38,* 72–78.

*Miller, R. B., Perry, F. L., & Cunningham, D. J. (1977). Differential forgetting of superordinate and subordinate information acquired from prose material. *Journal of Experimental Psychology, 69,* 730–735.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16,* 519–533.

Mosteller, F., & Hoaglin, D. C. (1991). Preliminary examination of data. In D. C. Hoaglin, R. Mosteller, & J. W. Tukey (Eds.), *Fundamentals of exploratory analysis of variance* (pp. 40–49). New York: Wiley.

Mumford, M. D., Weeks, J. L., Harding, F. D., & Fleishman, E. A. (1987). Measuring occupational difficulty: A construct validation against training criteria. *Journal of Applied Psychology, 72,* 578–587.

Naylor, J. C., & Briggs, G. E. (1961). *Long-term retention of learned skills: A review of the literature* (ASD TR 61–390). Columbus: Ohio State University, Laboratory of Aviation Psychology.

*Naylor, J. C., Briggs, G. E., & Reed, W. G. (1968). Task coherence, training time and retention interval effects of skill retention. *Journal of Applied Psychology, 52,* 386–393.

*Neumann, E., & Ammons, R. B. (1957). Acquisition and long-term retention of a simple serial perceptual-motor skill. *Journal of Experimental Psychology, 53,* 159–161.

Noe, R. A. (1986). Trainee's attributes and attitudes: Neglected influences on training effectiveness: Test of a model. *Personnel Psychology, 39,* 497–523.

*Norsworthy, N. (1912). Acquisition as related to retention. *Journal of Experimental Psychology, 3,* 214–218.

Peters, L. H., & O'Connor, E. J. (1980). Situational constraints and work outcomes: The influence of a frequently overlooked construct. *Academy of Management Review, 5,* 391–397.

*Postman, L., & Rau, L. (1957). Retention as a function of the method of measurement. *University of California Publication of Psychology, 8,* 217–270.

Proctor, R. W., & Dutta, A. (1995). *Skill acquisition and human performance.* Thousand Oaks, CA: Sage.

*Proctor, R. W., Reeve, T. G., Weeks, D. J., Dornier, L., & Van Zandt, T. (1991). Acquisition, retention, and transfer of response selection skill in choice reaction tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 497–506.

Prophet, W. W. (1976). *Long-term retention of flying skills: A review of the literature* (HumRRO Final Rep. No. 76–35). Alexandria, VA: Human Resources Research Organization (ADA036077).

*Purdy, B. J., & Lockhart, A. (1962). Retention and relearning of gross motor skills after long periods of no practice. *Research Quarterly, 33,* 265–272.

*Rabbitt, P., Cumming, G., & Vyas, S. (1979). Improvement, learning, and retention of skill at a visual search. *Quarterly Journal of Experimental Psychology, 31,* 441–459.

Regian, J. W., & Schneider, W. (1990). Assessment procedures for predicting and optimizing skill acquisition after extensive practice. In N. Frederickson, R. Glaser, A. Lesgold, & M. G. Shafto (Eds.), *Diagnostic monitoring of skill acquisition and knowledge acquisition* (pp. 297–323). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

*Reynolds, B., & Bilodeau, I. M. (1952). Acquisition and retention of three psychomotor tests as a function of distribution of practice during acquisition. *Journal of Experimental Psychology, 44,* 1–26.

*Richards, J. P. (1976). Interaction of position and conceptual level of adjunct questions on immediate and delayed retention of text. *Journal of Educational Psychology, 68,* 210–218.

Roehrig, W. C. (1964). Psychomotor task with perfect recall after fifty weeks of no practice. *Perceptual and Motor Skills, 19,* 547–550.

*Rose, A. M., Czarnolewski, M. Y., Gragg, F. E., Austin, S. H., & Ford, P. (1985). *Acquisition and retention of soldiering skills* (Final Rep. No. ADA1603364XSP). Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences.

Rubin-Rabson, G. (1939). Studies in the psychology of memorizing piano music: I. A comparison of the unilateral and coordinated approaches. *Journal of Educational Psychology, 30,* 321–345.

Rubin-Rabson, G. (1940a). Studies in the psychology of memorizing piano music: II. A comparison of massed and distributed practice. *Journal of Educational Psychology, 31,* 270–284.

Rubin-Rabson, G. (1940b). Studies in the psychology of memorizing piano music: III. A comparison of the whole and part approaches. *Journal of Educational Psychology, 31,* 460–476.

Rubin-Rabson, G. (1941a). Studies in the psychology of memorizing piano music: IV. The effect of incentive. *Journal of Educational Psychology, 32,* 45–54.

Rubin-Rabson, G. (1941b). Studies in the psychology of memorizing piano music: V. A comparison of pre-study periods of various lengths. *Journal of Educational Psychology, 32,* 101–112.

Rubin-Rabson, G. (1941c). Studies in the psychology of memorizing piano music: VI. A comparison of two forms of mental rehearsal and keyboard overlearning. *Journal of Educational Psychology, 32,* 593–602.

Rubin-Rabson, G. (1941d). Studies in the psychology of memorizing piano music: VII. A comparison of three degrees of overlearning. *Journal of Educational Psychology, 32,* 686–696.

Ryan, E. D. (1962). Retention of stabilometer and pursuit motor skills. *Research Quarterly, 33,* 593–598.

*Ryan, E. D. (1965). Retention of stabilometer performance over extended periods of time. *Research Quarterly, 36,* 46–51.

Ryan, E. D., & Simons, J. (1981). Cognitive demand, imagery, and frequency of mental rehearsal as factors influencing acquisition of motor skills. *Journal of Sport Psychology, 3,* 35–45.

Salas, E., Bowers, C., & Cannon-Bowers, J. A. (1995). Military team research: 10 years of progress. *Military Psychology, 7,* 55–75.

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

*Sassenrath, J. M., & Yonge, G. D. (1968). Delayed information feedback, feedback cues, retention set, and delayed retention. *Journal of Educational Psychology, 59,* 69–73.

Schab, F. R. (1990). Odors and the remembrance of things past. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 16,* 648–655.

Schendel, J. D., & Hagman, J. D. (1982). On sustaining procedural skills over a prolonged retention interval. *Journal of Applied Psychology, 67,* 605–610.

Schendel, J. D., Shields, J. L., & Katz, M. S. (1978). *Retention of motor skills: Review* (Technical Paper 313). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Schmidt, F. L., Law, K., Hunter, J. E., Rothstein, H. R., Pearlman, K., & McDaniel, M. (1993). Refinements in validity generalization methods: Implications for the situation specificity hypothesis. *Journal of Applied Psychology, 78,* 3–12.

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

*Semb, G. B., & Ellis, J. A., & Araujo, J. (1992). *Long term memory for different types of classroom knowledge* (Report No. AD–A255 235). San Diego, CA: Navy Personnel Research and Development Center.

Shields, J. L., Goldberg, S. L., & Dressel, J. D. (1979). *Retention of basic soldiering skills* (Research Rep. No. 1225). Alexandria, VA: U.S. Army Research Institute.

Shuell, T. J., & Keppel, G. (1970). Learning ability and retention. *Journal of Educational Psychology, 61,* 59–65.

*Slamecka, N. J., & McElree, B. (1983). Normal forgetting of verbal lists as a function of their degree of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 384–397.

Smith, J. F., & Matheny, W. G. (1976). *Continuation versus recurrent pilot training* (Report No. AFHRL–TR–76–4). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Smith, S. M., Glenberg, A. M., & Björk, R. A. (1978). Environmental context and human memory. *Memory and Cognition, 6,* 342–353.

*Sulin, R. A., & Dooling, D. J. (1974). Intrusion of a thematic idea in a retention of prose. *Journal of Experimental Psychology, 103,* 255–262.

*Swezey, R. W., Perez, R. S., & Allen, J. A. (1988). Effects of instructional delivery system and training parameter manipulations on electromechanical maintenance performance. *Human Factors, 30,* 751–762.

*Swezey, R. W., Perez, R. S., & Allen, J. A. (1991). Effects of instructional strategy and motion presentation conditions on the acquisition and transfer of electromechanical troubleshooting skill: Training theory, methods, technology [Special issue]. *Human Factors, 33,* 309–323.

Swezey, R. W., & Salas, E. (1992). Guidelines for use in team training development. In R. W. Swezey & E. Salas (Eds.), *Teams: Their training and performance* (pp. 219–245). Norwood, NJ: Ablex.

Swift, E. J. (1906). Memory of skillful movements. *Psychological Bulletin, 3,* 185–187.

Tannenbaum, S. I., & Yukl, G. (1992). Training and development in work organizations. *Annual Review of Psychology, 43,* 399–441.

Therrien, M. E. (1979, September). Evaluating empathy skill training for parents. *Social Work,* 417–419.

Towne, B. M. (1922). An individual curve of learning: A study in typewriting. *Journal of Experimental Psychology, 5,* 79–92.

Trumbo, D., Noble, M., Cross, K., & Ulrich, L. (1965). Task predictability in the organization, acquisition, and retention of tracking skill. *Journal of Experimental Psychology, 70,* 252–263.

Trumbo, D., Noble, M., & Swink, J. (1967). Secondary task interference in the performance of tracking tasks. *Journal of Experimental Psychology, 73,* 232–240.

*Trumbo, D., Ulrich, L., & Noble, M. (1965). Verbal coding and display coding in the acquisition and retention of tracking skills. *Journal of Applied Psychology, 49,* 368–375.

Tsai, C. (1924). A comparative study of retention curves for motor habits. *Comprehensive Psychology Monograph, 2*(Whole No. 11).

Tukey, J. W. (1960). A survey of sampling from contaminated distributions. In I. Olkin, J. G. Ghurye, W. Hoeffding, W. G. Madoo, & H. B. Mann (Eds.), *Contributions to probability and statistics* (pp. 448–485). Stanford, CA: Stanford University Press.

Tukey, J. W. (1977). *Exploratory data analysis.* Reading, MA: Addison-Wesley.

Tulving, E. (1983). *Elements of episodic memory.* New York: Oxford University Press.

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

Tulving, E., & Weisman, S. (1975). Relation between recognition and recognition failure of recallable words. *Bulletin of Psychonomic Society, 6,* 78–82.

*Underwood, B. J., & Keppel, G. (1963). Retention as a function of degree of learning and letter-sequence interference. *Psychological Monographs, 77*(Whole No. 567).

*Van Dusen, F., & Schlosberg, H. (1948). Further study of the retention of verbal and motor skills. *Journal of Experimental Psychology, 38,* 526–534.

*Vineberg, R. (1975). *A study of the retention skills and knowledge acquired in basic training* (HumRRO TR 75–10). Alexandria, VA: Human Resources Research Organization.

*Wetzel, S. K., Konoske, P. J., & Montague, W. E. (1983). *Estimating skill loss throughout a Navy technical training pipeline* (Final Rep. No. ADA1366368XSP). San Diego, CA: Navy Personnel Research and Development Center.

Wexley, K. N., & Latham, G. P. (1991). *Developing and training human resources in organizations.* New York: HarperCollins.

Wisher, R. A., Sabol, M. A., Hillel, H., & Kern, R. P. (1991). *Individual ready reserve (IRR) call-up: Skill decay* (ARI Research Rep. No. 1595). Alexandria, VA: U.S. Army Research Institute.

Wolf, F. M. (1986). *Meta-analysis: Quantitative methods for research synthesis.* Newbury Park, CA: Sage.

*Wright, R. H. (1973). *Retention of flying skills and refresher training requirements: Effects of non-flying and proficiency flying* (HumRRO Tech. Rep. No. 73–32). Alexandria, VA: Human Resources Research Organization.