UNIVERSITY OF OKLAHOMA

GRADUATE COLLLEGE

FACTORS INFLUENCING KNOWLEDGE AND SKILL DECAY IN

ORGANIZATIONAL TRAINING: A META-ANALYSIS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

XIAOQIAN WANG
Norman, Oklahoma
2010

FACTORS INFLUENCING KNOWLEDGE AND SKILL DECAY IN
ORGANIZATIONAL TRAINING: A META-ANALYSIS

A DISSERTATION APPROVED FOR THE

DEPARTMENT OF PSYCHOLOGY

BY

_____

Dr. Eric A. Day, Chair

_____

Dr. Mark Bolino

_____

Dr. Jorge Mendoza

_____

Dr. Lori Anderson Snyder

_____

Dr. Robert Terry

DEDICATION

This work is dedicated to my beloved parents, Jianhua Wu and Huichuan Wang, who

love and believe in me.

Acknowledgements

First and foremost, I would like to thank my advisor, Dr. Eric Day, for his guidance and support throughout the entire process as well as my years at the University of Oklahoma. I am very grateful that I was offered the opportunity to work with and learn from Eric, and it all started with my first conversation with Eric over lunch at SIOP in 2004. I would also like to express my gratitude towards my committee, Mark Bolino, Jorge Mendoza, Lori Anderson Snyder, and Robert Terry for their insights and continued support throughout. My sincere thanks will also go to the people who offered their generous help with this project: Vanessa Kowollik, Matthew Schuelke, Michael Hughes, and my undergraduate research assistants. Last but not the least, many thanks to my loving family and dear friends for your encouragement every step of the way and for sharing my exploration of life with me.

I would like to end this acknowledgement with a quote from a dear friend: "Someone once said that the happiest person is not the one who lives the longest life, but the one who lives through most moments. I think that is somewhat related to dreaming about what lies south of the border. Maybe it is easier for me to say that one should at least be open to the prospect of discovering what lies outside the small Siberian farm…The process of discovery teaches us, through experiences both good and bad, and makes us richer persons. But is it worth giving up a life filled with love for something that is…'probably' unattainable?"

Table of Contents

List of Tables

Summary of Multiple Regression Results

Meta-Analysis Results for Task Content by Retention Interval Analyses

List of Figures

Scree Plot for Sample-adjusted Meta-analytic Deviancy Statistics

Meta-analytic Results for Tasks Content by Retention Interval

Abstract

The current meta-analysis summarized existing organizationally-relevant training research on knowledge and skill decay. Results based on 111 independent effects retrieved from 35 manuscripts suggested an overall moderate decay effect ($\delta = -0.38$). The amount of decay was minimal for periods of nonuse less than one day ($\delta = -0.08$); decay was moderate to large for longer periods of nonuse (δs ranged from $= -0.24$ to -0.84), but there was no clear linear relationship between the amount of decay and length of nonuse. Decay was related to several methodological factors including the operationalization of acquisition, type of evaluation criteria, degree of training structure, and the use of post-training decay-prevention interventions. Decay was related to the combination of cognitive and physical task demands as well as task complexity. Tasks with moderate cognitive demands and minimal physical demands were associated with the greatest decay, whereas tasks with minimal cognitive demands but strong physical demands were associated with the least decay. Complex tasks were associated with modest levels of decay that were unlikely to be moderated by additional factors, but for simpler tasks decay effects were less robust and dependent upon other factors. Regression analysis indicated that decay was primarily related to the length of nonuse, amount of cognitive task demands, and the closed-/open-looped distinction. Longer periods of nonuse, greater cognitive demands, and closed-loop tasks were associated with greater decay. These results are discussed with respect to previous meta-analyses of decay effects. Both practical and theoretical implications are also discussed.

Factors Influencing Knowledge and Skill Decay in Organizational Training:

A Meta-analysis

Training is an essential endeavor in many organizations. Based on a 1998 report

by the American Society for Training and Development, an estimated $55.3 billion was

spent on formal training among U.S. organizations (Bassi & Van Buren, 1998). Such

spending increased from $110 billion in 2006 (Rivera & Paradise, 2006) up to about

$134.1 billion on employee learning and development last year (Paradise & Patel,

2009). However, it is commonly thought that only 10% of these dollars invested in

training result in "enduring behavioral change" (Wexley & Latham, 2002, p. 261).

These low returns on investment are costly to the organizations implementing training.

From an organization's perspective, maximizing the amount of post-training knowledge

and skills retained is directly linked to achieving a high return on investment.

Decay refers to the loss of trained or acquired knowledge and skills after a given

period of nonuse (Arthur, Bennett, Stanush, & McNelly, 1998). It should be noted that

decay and retention represent two sides of the same coin, as shown in the following

equation:

$$d = \frac{M_{T2} - M_{T1}}{\sigma_{pooled}}$$

$$\sigma_{pooled} = \sqrt{\frac{(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2}{n_1 + n_2}}$$

where d is the effect size, M_{T1} is the immediate posttest mean (i.e., at the end of a given

training program), M_{T2} is the delayed posttest mean (i.e., at some time after training and

the immediate posttest), and σ_{pooled} is the pooled standard deviation (Dunlap, Cortina, Vaslow, & Burke, 1996). A negative d would show decay after the retention interval (i.e., period of nonuse as indicated by the time between posttests), with larger effects indicating more decay (or less retention). A d with a non-negative value (i.e., $d = 0$ or positive values) indicates no decay, if not gain after training. Hence, both terminologies—decay and retention—will be used interchangeably throughout the rest of the manuscript.

With one exception (i.e., Arthur et al., 1998), existing reviews of decay and retention have been qualitative in nature. Both the qualitative and quantitative reviews found that decay increases as the retention interval extends. Specifically, Arthur et al.'s (1998) meta-analysis showed that the correlation between retention interval and corrected mean d was –0.51, indicating that decay increases as retention interval increases. In addition, several other factors, such as degrees of overlearning and certain task characteristics, were also found to influence decay after periods of nonuse in both the qualitative and quantitative reviews. The Arthur et al. (1998) meta-analysis differentiated methodological factors from task-related factors that influence decay. Results for the methodological factors showed that less decay for behavioral criteria ($d = -0.77$) than for learning criteria ($d = -1.04$); less decay when conditions of retrieval were similar to those during original learning ($d = -0.94$) than when those were different ($d = -2.07$); and a small difference in the amount of decay for studies that used recognition tests ($d = -0.85$) versus recall tests ($d = -0.96$). Due to limited data at the time, no conclusive statements were made by Arthur and colleagues regarding the influence of degrees of overlearning on decay. The findings of task-related factors

indicated more decay for open-looped tasks ($d = -1.04$) than for closed-looped tasks ($d = -0.71$); more decay for cognitive ($d = -1.15$) than for physical tasks ($d = -0.75$); and more decay for accuracy ($d = -1.00$) than for speed tasks ($d = -0.32$). Additionally, Arthur et al. suggested exploring in future research the effect of additional moderators, such as task complexity, distribution of practice, and decay-prevention interventions during retention intervals.

Arthur et al.'s (1998) initial efforts were not without limitations partially due to the limited number of empirical studies available at the time. Also the dichotomization of each of the task-related characteristics studied in their meta-analysis was rather coarse and it failed to reflect the multi-dimensional nature of those factors. The variable categories were found to covary and thus their moderator analyses involving coarse categorical variables were confounded. For example, physical tasks were more likely to be speeded than accuracy-based with an even distribution of open- versus closed-loop, whereas cognitive tasks were more likely to be accuracy-based than speeded and open-rather than closed-looped. In addition, Arthur et al. (1998) focused only on studies with a strict nonuse retention interval, while not addressing post-training factors, such as opportunity to practice or other decay-prevention interventions, which have been demonstrated to influence decay both in theoretical models and empirical studies (e.g., Baldwin & Ford, 1988; Gaudine & Saks, 2004).

The purpose of the present study was to conduct a meta-analysis of organizationally-relevant training research to determine the methodological and task factors most likely to influence the decay of trained knowledge and skills. Given that the Arthur et al. (1998) meta-analysis is now over 10 years old, an updated meta-

analysis seems appropriate. Additionally, to improve upon the limitations of Arthur's (1998) study, I (1) examined methodological factors previously not examined, (2) used a more sophisticated, multidimensional approach to coding task characteristics that involves coding dimensions along a 3-point scale (0-2), and (3) performed analyses that take into account the interaction and covariation among task-related characteristics as well as methodological factors. With respect to methodological factors, like Arthur et al. (1998), I focused on the length of retention interval, the degree of overlearning, condition of retrieval, and type of evaluation criteria. To expand upon Arthur et al. (1998), I also examined the influence of how acquisition was operationalized, degree of training structure, practice opportunities, and decay-prevention interventions. With respect to task-related characteristics, I focused on differences in decay as a function of task content and complexity. To expand upon Arthur et al. (1998), I added interpersonal demands to the examination of content beyond cognitive and physical demands. My focus on complexity expanded upon Arthur et al.'s (1998) closed- versus open-loop distinction.

<div align="center">Major Factors that Influence Decay</div>

Methodological Factors

Retention interval refers to the duration between the criterion assessment and the most recent training session (Ebbinghaus, 1885, 1913; Driskell, Willis, & Copper, 1992). Generally speaking, performance is negatively related to retention interval length, with longer retention intervals resulting in worse performance than shorter retention intervals due to the increased possibilities of interference and forgetting during the extended delay between acquisition and retrieval (Driskell et al., 1992; Farr, 1987).

Consistent with the findings in Arthur et al. (1998), I examined the following hypothesis.

H_1: There will be a negative relationship between length of retention interval and retention; the longer the retention interval, the more decay will be.

Initial acquisition. Initial knowledge and skill acquisition is a vital prerequisite of knowledge and skill retention (Arthur et al., 1998; Farr, 1987; Schmidt & Bjork, 1992), yet it is not equivalent to permanent or long lasting changes (Kraiger, 2003). Theoretically, there are three stages of cognitive skill acquisition. According to the Anderson's ACT* model (Anderson, 1987, 1996), these cognitive skill acquisition stages progress from acquiring *declarative knowledge* (memory-based), to *knowledge compilation* (sequencing steps required to perform a task), and eventually acquiring *procedural knowledge* (automated performance). Despite the argument for automaticity as indications for competent performance (Howell & Cooke, 1989), the distinction between these stages in empirical studies is rather arbitrary.

Depending on the purpose of a given training intervention, trainees are expected to achieve a specified level of performance (i.e., criterion-based training) or to spend a specific amount of time (i.e., duration-based training) in training (Adams & Hufford, 1962; Arthur, Day, Bennett, McNelly, & Jordan, 1997; Rohrer, Taylor, Pashler, Wixted, & Cepeda, 2005). Most researchers would use a measure of performance immediately after training as an operationalization of initial skill acquisition. Arthur et al. (1998) did not directly address operationalization of initial skill acquisition as a methodological moderator in their meta-analysis, but they did suggest that future research should pay more attention to initial acquisition as the prerequisite for retention. Without having

acquisition clearly defined or operationalized, discussion of decay is less meaningful. Therefore, the current meta-analysis examined if decay differed depending on how initial acquisition was operationalized.

Q_1: Will decay differ depending on whether the operationalization of initial

acquisition is criterion-based or duration-based?

Overlearning refers to deliberate learning and practice beyond the point of initial mastery, which can be either completing the first error-free trial or reaching a set criterion performance (Ebbinghaus, 1885, 1913). In other words, overlearning is relevant only in criterion-based training. In meta-analyses by Driskell, Willis, and Cooper (1992) and Arthur et al. (1998), *degree of overlearning* (DOV) was operationalized as follows:

$$\text{DOV} = \frac{\text{\% Learning in Overlearned Condition}}{\text{\% Learning in Overlearned Condition} + \text{\% Learning in Non-overlearned Condition}}$$

The meta-analysis on overlearning by Driskell et al. (1992) showed that greater overlearning led to less decay of newly learned cognitive (e.g., remembering verbal information) and physical skills (e.g., balancing on a stabilometer), and overlearning was more effective for retaining cognitive skills ($d = 0.75$) compared to physical skills ($d = 0.44$). Consistent with findings from Driskell et al. (1992), Arthur et al. (1998) hypothesized the positive relationship between overlearning and skill retention but did not find solid support for the hypothesis. Given the relatively weak effect based on 30 studies (17% of the entire data points) with a limited range of degree of overlearning, Arthur and colleagues (1998) cautioned readers of the interpretation regarding the minimum differences in decay as the degree of overlearning increased from 60% to

78.9%. It should also be noted that a large portion of studies included in both meta-analyses involved tasks, such as remembering verbal information or balancing on a stabilometer, that were simple and straightforward. For those tasks, it was easy to identify when the first error-free trail (i.e., initial mastery) occurred. However, the determination of initial mastery would be challenging for complex and dynamic tasks in the context of organizational training, which is the focus of the current meta-analysis. Nevertheless, based on the existing literature on the positive effect overlearning on facilitating retention, the following hypothesis was examined.

> H_2: Decay will decrease as the degree of overlearning increases in criterion-based training.

Condition of retrieval. According to the encoding specificity principle (Tulving, 1983; Tulving & Thompson, 1973) and the identical-elements theory (Thorndike & Woodworth, 1901), the likelihood of information retrieval increases as the similarity of retention testing and original learning context increases. Similarities between these two contexts provide cues and stimuli that facilitate information retrieval. Machin (2002) suggested that one should ensure that training incorporates procedures that are similar to those in the workplace, in order to maximize skill retrieval after training. Arthur et al. (1998) found that the amount of decay was significantly greater when the retrieval conditions were different from those of initial acquisition. Therefore, I examined the following hypothesis.

> H_3: Decay will be greater when the conditions between the original and retrieval assessment are different than when they are similar.

Types of evaluation criteria. Presently, there are two major taxonomies of training evaluation criteria. First, Kirkpatrick's typology (Kirkpatrick, 1959, 1976, 1996) differentiates reaction (i.e., trainees' affective reaction and perceived utility of training), learning (i.e., measures of knowledge and skill after training is completed), behavioral (i.e., on-the-job performance), and results (i.e., utility of training in terms of meeting organizations' objectives) as four criteria at different hierarchical levels. Although this typology is widely used by both researchers and practitioners, Kirkpatrick's typology has been criticized for its lack of theoretical grounding and over-simplification in treating each criterion level as unidimensional. Alliger and Janak (1989) also challenged three assumptions of Kirkpatrick's approach regarding the implicit hierarchical structure of four criteria at different levels and the assumed positive relationship among these criteria.

A more construct-centered taxonomy by Kraiger, Ford, and Salas (1993) categorizes training evaluation criteria into cognitive, skill-based, and affective outcomes. Although both taxonomies exhibit some conceptual overlap, they offer important distinctions. The Kraiger et al. taxonomy, which is more consistent with current learning theory, provides a more theoretically viable framework in evaluating training by distinguishing between different types of cognitive, skill-based, and affective outcomes. For example, the Kraiger et al. taxonomy extended the reaction criteria by proposing a category called affective outcomes that include both attitudinal (e.g., respect diversity at workplace) and motivational (e.g., changes in self-efficacy) outcomes, rather than equating attitudinal outcomes to reactions (c.f., Kirkpatrick, 1959, 1976, 1996). Another valuable addition by Kraiger and colleagues is the inclusion of

knowledge organization and cognitive strategies besides verbal knowledge (similar to learning in the Kirkpatrick typology) under cognitive outcomes. Last, based on theories of skill development (e.g., Anderson, 1996), the Kraiger et al. taxonomy categorized skill-based outcomes into proceduralization (i.e., reproduction of trained skills), composition/adaptability (i.e., application of trained skills in a new context or with new contingencies), and automaticity (i.e., performing the task without conscious monitoring).

Only learning and behavior criteria under the Kirkpatrick's typology were used by Arthur and colleagues in their skill decay meta-analysis (Arthur et al., 1998). They found that the amount of decay was lower for behavioral ($d = -0.77$) than for learning criteria ($d = -1.04$). The current meta-analysis used the more construct-centered Kraiger et al. taxonomy, and focused on cognitive and skill-based outcomes. Although cognitive and skill-based criteria under the Kraiger et al. taxonomy are similar to learning and behavioral criteria under the Kirkpatrick's typology, I proposed the following research questions rather than framing them as hypotheses due to the lack of past research in this regard.

Q_2: Will rates of decay differ for cognitive criteria compared to skill-based criteria?

Q_3: Will rates of decay differ for different types of cognitive criteria?

Q_4: Will rates of decay differ for different types of skill criteria?

Structure of the training, as an important variable to consider in training, refers to the extent to which instructors provide guidance, clarification, feedback and objectives for trainees, and the extent to which instructors exert control over how the

materials are delivered, as well as trainees' learning process (Campbell & Kuncel, 2001). As the degree to which these elements in a training program increases, the structure of the training increases (Freebody & Tirre, 1985; Snow, 1989). In training programs with high structure, instructors are in charge of how instructional tasks are broken down into small units and delivered using different training activities. On the contrary, in training programs with low structure, trainees take a more proactive role and rely more on themselves, rather than on instructors in the learning process. Some research has shown that training with low structure, such as guided exploratory training, is beneficial in leading to better strategic knowledge, higher performance and better transfer in various tasks (e.g., Dormann & Frese, 1994; McDaniel & Schlager, 1990). However, it has also been shown that trainees are likely to be at a loss when offered with too much freedom due to cognitive overload as a result of being overwhelmed with making all the instruction decisions, resulting in a failure to focus proper attention on important content in training (Mayer, 2004; DeRouin, Fritzsche, & Salas, 2004; Salas, Cannon-Bowers, Rhodenizer, & Bowers, 1999). It is commonly recommended to provide instructional messages that guide trainees' exploration during training so as to prevent them from deviating from the productive learning process pre-designed by instructors (Mayer, 2004). For example, active learning, as a training strategy, not only offers learner control to trainees but it also involves formal training design elements to guide the cognitive, motivational, and emotional learning processes, which consequently facilitate self-regulated learning and transfer (Bell & Kozlowski, 2008; Bransford, Brown, & Cocking, 1999). Therefore, to examine how structure of training would influence decay, the following research question was examined:

Q_5: Will there be a relationship between training structure and decay? If

so, what is the nature of the relationship?

Practice opportunities refer to the extent to which trainees are provided with or actively obtain work experience relevant to the trained tasks after training (Ford, Quiñones, Sego, & Sorra, 1992). Quiñones, Ford, Sego, & Smith, (1996) found that the more opportunities trainees have to apply acquired knowledge and skills, the more likely trainees can replicate such knowledge and skills in the actual work environment. Although the opportunity to practice is considered to be an important factor influencing decay, it was not examined by Arthur et al. (1998) because their meta-analysis only investigated studies involving nonuse intervals. Therefore, the following hypothesis was examined:

H_4: Opportunities to practice will be associated with lower levels of decay.

Decay-prevention intervention. Post-training decay-prevention interventions intend to promote effective transfer of training to the job environment by helping trainees to set transfer goals, creating a supportive transfer climate, monitoring posttraining performance, and providing relapse prevention training (Foxon, 1993, 1994; Goldstein & Ford, 2002; Machin, 2002; Salas et al., 1999). For example, goal-setting and self-management are the most commonly used post-training decay-prevention interventions. *Goal-setting* posits that difficulty and specificity are the two components of goals that relate to performance (Locke & Latham, 1990). *Self-management training* is related to *relapse prevention* that was originally applied in addictive behaviors prevention (Marx, 1982). Self-management training is a cognitive-behavioral strategy structured in order to encourage trainees to identify obstacles and

strategies to overcome them. As supported by findings from relevant primary studies and qualitative reviews, both goal-setting and self-management interventions have been found to be effective in promoting retention in training on negotiation, customer service, and supervisory skills (Gist, Stevens, & Bavetta, 1991; Hutchins & Burke, 2006; Noe, Sears, & Fullenkamp, 1990; Richman-Hirch, 2001; Tews & Tracey, 2008). Therefore, the following hypothesis was examined:

H_5: The use of post-training decay-prevention interventions will be associated with lower levels of decay.

Task-related Factors

Task content concerns the nature of a task that trainees are expected to learn in training. Depending on the requirements held for trainees to learn, the task content can fall into the following three categories: cognitive, physical, and interpersonal tasks (Farina & Wheaton, 1973; Fleishman & Quaintance, 1984; Goldstein & Ford, 2002). Cognitive tasks involve information processing, problem solving, sensemaking, idea generation, and decision making. Physical tasks require the use of the musculoskeletal system to perform a range of physical movements. Interpersonal tasks relate to interacting with others including co-workers, clients and customers. However, the content of many tasks involves some mix or combination of cognitive, physical or interpersonal components. For example, Space Fortress is an experimental game designed to simulate a complex and dynamic aviation environment (Mané & Donchin, 1989). Space Fortress not only has a high cognitive load in information processing and decision-making, but it also involves difficult joystick and mouse controls. Rather than

forcing Space Fortress into only one task content type, it can be categorized as both a cognitive and physical task (c.f., Arthur, Bennett, Edens, & Bell, 2003).

Although previous research has shown that training effectiveness is different depending upon the content of the training (Aguinis & Kraiger, 2009; Wexley & Latham, 2002), such investigations have not specifically addressed training effectiveness with respect to decay. Moreover, Arthur et al.'s (1998) meta-analysis only compared decay by treating task content as a categorical variable: either cognitive or physical. Their results showed an overall greater amount of decay for cognitive tasks ($d = -1.15$) than for physical tasks ($d = -0.75$). In the present study, instead of making a categorical distinction, task content was treated as a multidimensional construct and scored on a three-point scale (0 = low [minimally applicable/involved]; 1 = moderate [somewhat applicable/involved]; 2 = high [very applicable/involved]) for each of the three content dimensions: cognitive, physical, and interpersonal. Table 1 shows a list of example tasks scored in this manner. This coding scheme allowed me to examine decay as a function of each type of content demand as well as in combination.

Arthur et al. (1998)'s dichotomization of cognitive versus physical tasks can be represented as tasks with high cognitive demands and low physical demands versus tasks with low cognitive demands and high physical demands under the current coding scheme. Therefore, I examined the following hypothesis.

H_6: There will be more decay for tasks with high cognitive demands and low physical demands than for tasks with low cognitive demands and high physical demands.

Furthermore, specific dimensions and different combination of specific content demands are likely to play a more or less important role in influencing decay. The multidimensional coding scheme of task content allowed me to examine the following research questions:

Q_6: Will decay be related to the strength of each of the content demands?

Q_7: Will decay differ for tasks that have greater combined demands (sum of cognitive, physical, and interpersonal) than for tasks with lower combined demands?

Task complexity is a function of objective characteristics of tasks and has been found to be an important determinant of individual's task performance (e.g., Haerem & Rau, 2007; Wood, 1986). The most commonly used framework of task complexity, with a focus on individual task performance, was articulated by Wood (1986). Under this taxonomy, task complexity is divided into three components: *component complexity* (the number of distinct acts and distinct information cues or information elements that are needed to perform the task), *coordinative complexity* (the form and strength of the relationships between the components of the task and the sequencing of the inputs), and *dynamic complexity* (the need to adapt to potential changes in the means-ends hierarchy during the performance of the task). Task complexity, as defined using Wood's taxonomy, is often studied as a moderator in meta-analysis. For example, Wood, Mento, and Locke (1987) showed that the magnitude of goal-setting effect on performance was moderated by task complexity, such that the magnitude of goal effects was greater on simple tasks (e.g., recall task) than on complex tasks (e.g., business game simulation). Chen, Casper, and Cortina (2001) found that self-efficacy partially mediated the

relationship of cognitive ability and conscientiousness with performance on simple

tasks, but not on complex tasks. In addition, the relationship of cognitive ability and

conscientiousness with performance became stronger as task complexity increased,

while self-efficacy was not related to complex task performance over and above

cognitive ability and conscientiousness (Chen et al., 2001).

A similar taxonomy used to categorize tasks was used in the meta-analysis of

the distributed practice effect by Donovan and Rodosevich (1999). Based on results

from cluster analysis of task characteristics, Donovan and Rodosevich categorized tasks

into different types along the following three dimensions. First, *overall task complexity*

was operationalized as the number of distinctive behaviors required, number of choices

needed, and degree of uncertainty involved to complete the task. This dimension was

adapted from the aggregated task complexity under the Wood (1986) taxonomy. The

other two dimensions: *mental requirement of the task* and *physical requirement of the*

task were unique dimensions that resulted from cluster analysis. These two dimensions

were not included in the operationalization of task complexity in the current meta-

analysis because they were captured under the aforementioned task content factor (i.e.,

cognitive, physical and interpersonal tasks). Results from their meta-analysis on

practice effects indicated that the superiority of spaced practice over massed practice

diminished as complexity of tasks on all three dimensions increased from low to high (d

$= 0.97$ to $d = 0.07$).

The current meta-analysis adopted the operationalization of task complexity

proposed by Wood (1986), and supplemented it with Donovan and Rodosevich (1999)'s

operationalization of overall task complexity as well as characteristics that differentiate

open-looped from closed-looped tasks. The scheme was as follows: except for whether a task is open-looped or not which was coded as yes or no, task complexity was scored on a 3-point scale (0 = low [minimally applicable/involved]; 1 = moderate [somewhat applicable/involved]; 2 = high [very applicable/involved]) for each of the three dimensions (i.e., discretion, dynamic complexity, and component complexity). Table 2 shows a list of example tasks scored in this manner. This coding scheme allowed me to examine decay as a function of each dimension of task complexity as well as in combination.

First, an open-looped task refers to a task without a clear and distinct end of the task, which is contingent on time, and requires constant monitoring of the discrepancies between the current and desired state. On the other hand, a closed-looped task is one with an unambiguous ending that is not contingent on time (Naylor & Briggs, 1961). For example, monitoring a nuclear reactor is an open-looped task that requires ongoing comparison between current and desired state of the reactor. The end of such tasks is usually marked by specific time limits. However, package delivery as a task is closed-looped because the task is complete when the package reaches the addressees' hands. Closed-looped tasks often involve discrete responses that have a definite beginning and end, whereas open-looped tasks typically involve continuous responses that are repeated or constant monitoring and do not have a definite beginning or end. Results for whether closed-looped or open-looped tasks are better retained are mixed (e.g., Arthur et al., 1998 versus Farr, 1987).

Results from primary studies and narrative reviews on closed-looped and open-looped tasks suggested that open-looped tasks were more resistant to decay than closed-

looped tasks. Compared to closed-looped tasks, open-looped tasks are more coherent and continuous in nature, so they are more natural and meaningful to people (Naylor & Briggs, 1961). Arthur et al. (1998) hypothesized that closed-looped tasks would decay faster than open-looped tasks because the latter involve continuous responses and are more coherent in nature. However, their findings indicated the opposite: open-looped tasks generally decayed more ($d = -1.06$) than closed-looped tasks ($d = -0.71$), which was not in line with that from primary studies and narrative reviews on closed-looped and open-looped tasks. This inconsistent finding may be partially due to the covariation among three types of task characteristic moderators (closed-/open-looped, physical/cognitive, and natural/artificial). Arthur and colleagues (1998) found that physical tasks were more likely to be speeded than accuracy-based with an even distribution of open- versus closed-loop, whereas cognitive tasks were more likely to be accuracy-based than speeded and open- rather than closed-looped.

Second, *discretion* refers to the number of choices or approaches that can be adopted to perform the task and reach the goal. This is adapted from coordinative complexity dimension under Wood's taxonomy (Wood, 1986) and the sub-dimensions of overall complexity in Donovan and Rodosevich (1999). Third, *dynamic complexity,* as specified under Wood's (1986) taxonomy, refers to the extent of the dynamic nature of the task environment. Task elements are dynamic in cases where inconsistent information exists or there is inconsistency in information processing and decision execution. In other words, a task with dynamic elements is characterized by the extent of not knowing what will happen or what to expect, and the requirement of constantly paying attention to different components in the task environment. Last, *component*

complexity, as specified by Wood (1986), refers to the total number of distinct

behaviors required to execute the task, and the total number of distinct information cues

that must be processed in performing those actions.

As the task complexity increases, the knowledge and skill requirements for

performing the task also increase. The characteristics of complex tasks are inherently

more meaningful to individuals. Rather than performing rote tasks, individuals are

motivated to engage in deeper and more elaborative processing during training, and

consequently achieve better retention given the nature of complex tasks (Arthur et al.,

1998; Craik & Lockhart, 1972; Schmidt & Bjork, 1992; Wexley & Latham, 2002).

Therefore, the following hypothesis was examined:

H_7: Decay will be smaller for more complex tasks than for less complex

tasks.

Covariation among Moderators

Due to limited number of data points available, Arthur et al. (1998) were not

able to conduct hierarchical moderator analysis despite the covariation they found

among some task-related and methodological factors. To address this question, the

current meta-analysis took three approaches to examine the independent effect of

moderators and covariation among moderators: (1) conducting hierarchical moderator

analysis when *k* sizes were five or bigger per cell to ensure a meaningful interpretation,

(2) examining the correlations among moderators, as well as between moderators and

corrected mean effect sizes, and (3) examining the independent effects of moderators by

regressing effect sizes on moderators.

Methods

Literature Search

An extensive literature search was conducted, both electronically and manually, to identify empirical studies that have investigated skill retention and/or decay up to the year 2008. The electronic search covered nine computer databases (*Academic Search Elite, Business Source Elite, Defense Technical Information Center, Dissertation Abstract, Econlit, Educational Research Information Center, Government Printing Office, PsychINFO,* and *SocINDEX*). The electronic search was supplemented by a manual search of reference list of recent training literature reviews, as well as SIOP conference presentations (Aguinis & Kraiger, 2009; Alliger & Janak, 1989; Alliger, Tannenbaum, Bennett, Traver, & Shotland, 1997; Arthur et al., 1998; Arthur et al., 2003; Colquitt, LePine, & Noe, 2000). The following key words were used: *knowledge/skill retention, knowledge/skill maintenance, knowledge/skill perishability, knowledge/skill deterioration, knowledge/skill decay, knowledge/skill degradation, knowledge/skill acquisition, training evaluation, training efficiency, training effectiveness,* and *training*. The initial search resulted in approximately 10,459 English language citations, out of which 38 articles were retained based on the inclusion criteria. The final dataset was comprised of the following sources: 29 articles from journal articles, 4 from dissertations, 4 from technical reports, 1 unpublished but submitted manuscript.

Inclusion Criteria

A set of decision rules were used to determine whether to retain the studies given the purpose of the current meta-analysis. First, to be included in the meta-

analysis, a study must have investigated the effectiveness of an organizational training program over time. The study can either have occurred in a field or laboratory setting. Laboratory studies were included if it was judged that the author intended to generalize the findings to organizationally relevant settings. Studies published in educational journals were excluded unless the researchers specifically sought to generalize their findings to organizational settings. In addition, experiments using simple tasks, such as memorizing nonsense syllables (e.g., Mandler & Heineman, 1956) or drawing lines blindfolded (e.g., Trowbridge & Cason, 1932), were not included because findings from such studies are less likely to have implications in organizational contexts. Second, trainees had to be 18 years and older. Third, studies must have clearly indicated that the training was completed either because trainees reached training goals or underwent a specific amount of time as required by the training program. Fourth, the study had to report enough information about the training (e.g., nature, protocol, task trained) for any reasonable judgments about the training program could be made. Fifth, for a criterion to be considered as assessing retention over time, the same measure had to be administered both immediately (i.e., within one hour) upon completion of training and some time later. The time interval between immediate and delayed post test had to be greater than the time interval between the end of training and immediate post test. Sixth, studies had to report important statistics that allow for the computation of d. In other words, studies had to report sample sizes along with means and standard deviations for both the immediate and delayed measure of a criterion. Effect sizes can also be calculated if dependent t-statistics are reported along with the correlation between the immediate and

delayed posttests (Dunlap et al., 1996). Last, the current meta-analysis only focused on individual-level effects.

Data Set

    Nonindependence. Based on the inclusion criteria, an initial dataset consisted of 209 data points (*ds*) from 38 manuscripts. But not all effect sizes were independent if they were computed from the same group of participants. In the context of meta-analysis, nonindependence can be problematic because it could reduce the observed variability of effect sizes, artificially inflate sample sizes and effect sizes, and overweight the contributions of studies with multiple nonindependent data points (Arthur, Bennett, & Huffcutt, 2001). For analysis examining overall effects (e.g., retention interval and decay), data points were averaged across all types of criteria if they were from the same group of participants under the same manipulation in the same experiment. However, for moderator analyses (e.g., whether decay differed for cognitive versus skill criterion), under the same manipulation in the same experiment, data points associated with cognitive criterion and those with skill criterion were consider independent if that experiment had both cognitive and skill criteria. Consequently, nonindependent data points were averaged and resulted in 114 independent data points from 38 manuscripts. Out of the 114 independent data points, three studies from journal articles reported post-training intervention and they were excluded from the final dataset except for addressing Hypothesis 5, which was the use of post training decay-prevention interventions would be associated with lower levels of decay. Therefore, the final dataset used to address the majority of the hypotheses and research questions consisted of 111 independent data points from 35 manuscripts. The

sources of these data points were as follows: 66.7% studies from journal articles, 11.7% from dissertations, 20.7% from technical reports, and 0.9% unpublished but submitted manuscript.

Outliers. As in empirical studies, the presence of outliers poses a concern for meta-analyses. Huffcutt and Arthur (1995) developed and suggested the sample-adjusted meta-analytic deviancy (*SAMD*) statistic to detect outliers. The *SAMD* procedure "compares the value of each study effect size to the mean sample-weighted coefficient computed without that coefficient in the analysis, then adjusts that difference for the sample size of the study" (Arthur et al., 2001, p. 119). A *SAMD* statistic was calculated for each data point in the dataset and each of the absolute *SAMD* values was ranked and plotted in order from highest to lowest.

SAMD statistics were computed for each of the 111 independent data points, resulting in a mean *SAMD* of –0.08 (*SD* = 1.82). The resulting *SAMD* scree plot is presented in Figure 1, suggesting a total of five outliers. However, these five studies, which constituted 4.5% of the 111 effect sizes in the dataset, were retained because they were originally included in Arthur et al. (1998) meta-analysis. One of these studies was from a journal article by Reynolds and Bilodeau (1952); two were from a technical report by Thompson, Morey, Smith, and Osborne (1981); and two were from a technical report by Grimsley (1969).

Given one of the objectives of the current meta-analysis was to provide an update of Arthur et al. (1998)'s meta-analysis, efforts were made to extract all 53 articles coded by Arthur and colleagues. Out of all those articles, 43 hard copies were found, 18 of which failed to meet the current inclusion criteria. With the remaining 25

articles that were not excluded, only eight were kept in the current meta-analysis
(including the three articles that showed up as potential outliers in the scree plot)
because the rest did not report relevant statistics to calculate effect sizes.

Description of Methodological Factors

    Retention intervals. The retention interval (i.e., number of hours) between the
end of training and collection of the criterion data was coded. Retention interval for
criteria measured immediately upon completion of training was code as the zero hour,
and the lag between training and measures at a later time was recorded in hours as
reported in the primary studies. In order to compare results related to retention intervals
from the current meta-analysis with those from Arthur et al. (1998), the same
categorization scheme was used to break number of hours into different retention
interval categories: (1) less than 1 day; (2) greater than or equal to 1 day, but less than
or equal to 7 days; (3) greater than 7 days , but less than or equal to 14 days; (4) greater
than 14 days, but less than or equal to 28 days; (5) greater than 28 days, but less than or
equal to 90 days; and (6) greater than 90 days, but less than or equal to 180 days (see
Table 5). A total of eight retention interval categories were in Arthur et al. (1998) meta-
analysis. However, primary studies included in the current meta-analysis did not have
retention intervals that would fall under the last two categories under their
categorization scheme (i.e., greater than 180 days, but less than or equal to 365 days; or
greater than 365 days).

    Operationalization of initial acquisition. Each training program was categorized
as either being criterion-based or duration-based.

Overlearning. The degree of overlearning was calculated only for studies that were coded as criterion-based, using the equation discussed earlier in this document. For example, if reaching the criterion level requires 5 trials and the overlearning condition involves 5 additional trials, then the degree of overlearning is 100%. A degree of overlearning of zero indicates no overlearning occurred.

Condition of retrieval. The context of immediate and delayed assessment was coded as being the same or different based on the description provided in primary studies.

Types of evaluation criteria. Dependent variables in primary studies were hierarchically classified as different types of criteria under either cognitive- or skill-based outcomes based on Kraiger et al.'s taxonomy (1993) of evaluation criteria. Specifically, the cognitive criteria were further broken down to declarative knowledge, procedural knowledge, strategic knowledge, knowledge organization, and cognitive strategies criteria. Skill criteria were coded as either proceduralized skill or adaptive skill.

Training structure. The degree of structure was a composite of six dimensions operationalized as follows: (1) instructor controlled activities (i.e., instructor determines the content, sequence and timing of training activities), (2) frequency of instructional message (i.e., frequency of explicit guidance provided by instructor regarding training content and approaches needed to learn the content), (3) frequency of clarification of material (i.e., how frequent instructor extends beyond the basics through additional clarifying information) , (4) individual assistance from the instructor (i.e., actual personalized hands-on assistance provided by instructor by providing advice or

demonstration), (5) break-down of training into modules (i.e., training is broken into different components or modules to maintain trainees' attention and focus), and (6) provision of detailed objectives for trainees (i.e., explicit information provided to trainees regarding what they are expected to know or do by the end of a training activity or module). All aforementioned dimensions were coded on a 3-point scale (0 = little if any—involved less than 10% of training time; 1 = moderate degree—involved more than 10% but less than 50% of training time; 2 = large degree—involved more than 50% of training time) with a higher number indicating a more structured training. Although scores could range from 0 to 12, the distribution of overall structure scores in the current meta-analysis as shown in Table 3 ranged from 0 to 9. Thus, none of the data points in this meta-analysis involved a sample that underwent highly structured (i.e., guided) training. Based on the frequency distribution shown in Table 3, the overall structure scores was trichotomized for moderator analyses purpose as follows: low structure = 0—2 (k = 38), medium structure = 3—4 (k = 58), and high structure = 5—9 (k = 15).

Practice opportunities. During the retention interval, information about whether trainees were constrained to no practice was coded as yes or no.

Decay-prevention intervention. During the retention interval, information about whether there was any decay-prevention intervention (e.g., refresher training) was coded as yes or no.

Description of Task-related Factors

Task content. Three types of task content (i.e., cognitive, physical, and interpersonal) were coded. Each of the task type was coded using a 3-point scale (0 =

low [minimally applicable/involved]; 1 = moderate [somewhat applicable/involved]; 2 = high [very applicable/involved]) to indicate the nature of the task. For example, training on negotiation skills would be coded as 1 for cognitive, 0 for physical, and 2 for interpersonal.

Task complexity. Open-looped task was coded as "Yes" and closed-looped task was coded as "No". Discretion, dynamic complexity, and component complexity were coded using a 3-point scale (0 = low [minimally applicable/involved]; 1 = moderate [somewhat applicable/involved]; 2 = high [very applicable/involved]), with a higher number indicating higher levels of complexity.

Coding Accuracy and Interrater Agreement

Having received 20 hours of training using a training manual developed for the current meta-analysis, three graduate students coded the data reported in the current meta-analysis. All coders were assigned a common set of nine articles used to assess interrater agreement by comparing the values and categorization assigned by each coder for all variables of concern (Table 4). The levels of agreement was generally high, with a mean overall agreement of 97.01% (SD = 5.55). Discrepancies and disagreements related to this set of nine articles were resolved through consensus meetings involving all three graduate student coders.

Computation of Effect Sizes

The current meta-analysis used the effect size statistic (d) as the common metric to aggregate the effects across all studies (Hunter & Schmidt, 1990). The following equation was used to calculate d:

$$d = \frac{M_{T2} - M_{T1}}{\sigma_{pooled}}$$

$$\sigma_{pooled} = \sqrt{\frac{(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2}{n_1 + n_2}}$$

where M_{T1} is the immediately posttest mean, M_{T2} is the delayed posttest mean, σ_1 is the immediately posttest standard deviation, σ_2 is the delayed posttest standard deviation, n_1 is the immediately posttest sample size, n_2 is the delayed posttest sample size, and σ_{pooled} is the pooled standard deviation (Dunlap et al., 1996). A negative d in the current meta-analysis would suggest decay from immediate posttest to delayed posttest; d close to zero would indicate no decay, and a positive d would suggest not only retention but improvement over the retention interval.

Studies with larger sample size are likely to have smaller sampling errors, therefore Hunter and Schmidt (1990) recommended assigning more weight to effect sizes associated with bigger sample sizes in calculating overall mean d using the following equation:

$$\overline{d} = \frac{\Sigma dini}{Nt}$$

where $\overline{d}$ is the overall 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. In addition, the sample sizes in the current meta-analysis were uneven across studies as they varied from 4 to 119. To correct for the attenuating effect of unequal or unbalanced sample sizes, a bias multiplier (denoted as "A") was used to calculate corrected mean $\overline{d}$ as suggested by Hunter and Schmidt (1990). The bias multiplier was calculated as:

$$A = 1 + \left(\frac{0.75}{\overline{N} - 3}\right)$$

where N is the average sample size across studies. Then the corrected mean

$\overline{d}$ (detonated as δ) and corrected standard deviation of the population effect sizes

(detonated as SD_δ) are calculated using the following equations:

$$\delta = \frac{\overline{d}}{A}$$

$$SD_\delta = \frac{\sqrt{Var\delta}}{A}$$

where δ is the corrected mean effect sizes, SD_δ is the corrected standard deviation, and

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

Confidence interval was the effects of sampling error remains in uncorrected,

sample-size weighted mean effect size, with narrower confidence interval suggesting

higher accuracy. In other words, if another set of studies were drawn from the

population to conduct a new meta-analysis on the same topic, the mean effect sizes

found are likely to be within the range of values identified by the confidence interval

(Arthur et al., 2001; Hunter & Schmidt, 2004; Whitener, 1990). In the current meta-

analysis, a 95% confidence interval (CI) for the correct mean d was calculated as

follows:

$$CI_{95\%} = \overline{d} \pm (1.96 * SD_\delta)$$

Moderator Analyses. To assess the impact of each proposed factor on decay,

studies were separated into different subsets based on specific levels of the factor. An

overall and a subset mean effect size was calculated for each factor. For the current

meta-analysis, a factor was considered a meaningful moderator if (1) the difference

between level effects (respective δs) found was equal to or greater than 0.20, which is Cohen's (1992) standard for a small effect size; and (2) the confidence intervals for each effect did not substantially overlap. Furthermore, a variable is also thought to be having a moderating effect if the percent of variance explained by sampling error increases (i.e., the residual variances decreases) after separating a set of studies based on levels of the supposed moderator factor.

Results

As shown in Table 6, across 111 independent effects, there was decay since the completing of formal training ($\delta = -0.38$). However, the 95% confidence interval was rather large, covering a range from strong levels of decay ($\delta = -1.50$) to moderately strong levels of retention ($\delta = 0.75$). Furthermore, there was a small amount of variance accounted for by sampling error (30.85%), suggesting that the investigation of moderators is warranted.

Methodological Factors

Retention interval. The first hypothesis involved examining the effect of the length of nonuse retention interval on the amount of decay. Specifically, it was hypothesized that length of retention intervals would be negatively related to the amount of retention. In support of Hypothesis 1, the correlation between retention interval and corrected mean d (δ) was -0.58 ($p > .05$), indicating longer intervals were associated with more decay. Although this correlation was not significant given the limited six data points (i.e., six retention interval categories), the direction of the relationship was consistent with Hypothesis 1: longer retention intervals were associated with more decay. Results from Table 6 indicate that there was almost no

decay ($\delta = -0.08$) when the retention interval was less than one day. Beyond one day, higher levels of decay were found for Retention Intervals 2 through 6 compared to Retention Interval 1. However, there was not a linear trend for the amount of decay as retention interval extended. For instance, Retention Interval 5 (i.e., greater than 28 days, but less than or equal to 90 days) yielded less decay than Retention Intervals 2, 3, and 4. Overall, the results in Table 6 offered mixed support for Hypothesis 1. The shortest retention interval was associated with less decay than any other retention intervals, but there was no clear pattern of greater decay as retention interval increased beyond one day. However, the relatively small amount of variance accounted for by sampling error for Retention Intervals 2, 4 and 5 suggests the presence of additional moderators.

Further moderator analyses were conducted by collapsing across all retention intervals due to the small k sizes for most of the retention intervals. In addition, to ensure reasonable meaningful interpretations, hierarchical moderator analyses were only conducted when there were at least five independent effects (i.e., $k \geq 5$) that could be meta-analyzed for any given level of a moderator. However, if there was a specific hypothesis proposed, I conducted moderator analysis regardless of whether there were at least five independent effects.

Initial acquisition. Research Question 1 asked if decay would differ depending on whether the operationalization of initial acquisition is criterion-based or duration-based. Based on analyses on the limited seven studies that operationalized initial acquisition as criterion-based, the results showed there was more decay for criterion-based training ($\delta = -1.43$) than for duration-based training ($\delta = -0.35$). The absolute difference between these two effects was greater than 0.20, and the confidence intervals

for each effect did not substantially overlap (Table 7). Therefore, there was a meaningful difference in decay depending on whether the operationalization of initial acquisition is criterion-based or duration-based. However, the percent of variance accounted for by sampling error was relatively small for both effects, suggesting the presence of additional moderators.

Overlearning. Hypothesis 2 stated that less decay would be associated with increases in the degree of overlearning in criterion-based training. However, there were only seven studies ($N = 52$) that were criterion-based training, out of which no information regarding degree of overlearning was reported. Therefore, Hypothesis 2 could not be tested in the current meta-analysis.

Condition of retrieval. Hypothesis 3 stated that decay would be greater when the conditions between the original and retrieval assessment are different than when they are similar. As shown in Table 7, the results did not support Hypothesis 3. The amount of decay did not matter if the conditions between the original and retrieval assessment were different ($\delta = -0.38$) or similar ($\delta = -0.37$). However, it should be noted that the current finding was based on three data points that reported retrieval conditions being different from the original assessment conditions.

Criterion type. Three research questions were proposed regarding the rate of decay and different types of criteria as specified in the Kraiger et al. (1993) taxonomy. Research Question 2 asked whether rates of decay would differ for cognitive criteria compared to skill-based criteria. To address Research Question 2, all specific types of cognitive and skill criteria were collapsed, respectively. As shown in Table 7, the results indicated greater decay for cognitive criteria ($\delta = -0.60$) compared to skill criteria ($\delta = -$

0.25). It should also be noted that the range of the confidence interval associated with skill criteria (–1.54, 1.04) was larger than that with cognitive criteria (–1.07, –0.12). Furthermore, the confidence interval associated with skill criteria suggested the possibility of no decay or even improvement after a period of nonuse, which the same conclusion could not be drawn regarding cognitive criterion. Therefore, there was more decay for cognitive criteria compared to skill-based criteria.

Research Question 3 asked whether rates of decay would differ for different types of cognitive criteria. Under the *a priori* decision of excluding moderator analysis with less than five data points, a comparison was only made between two specific types of cognitive criteria: declarative knowledge versus declarative and procedural knowledge. It should be noted that no studies were available that involved a pure procedural knowledge criterion. Results showed less decay for declarative knowledge (δ = –0.46) than for a criterion assessing both declarative and procedural knowledge (δ = –0.93), and their confidence intervals did not overlap. The percent of variance accounted for by sampling error not only increased after separating cognitive criteria into these two specific cognitive criterion types, but also the high percent of variance accounted for by sampling error suggested the low likelihood of additional moderators that would make a meaningful difference. Research Question 4 asked whether rates of decay would differ for different types of skill criteria. Similarly, *a priori* decision rule was applied, and skill criteria were broken down into two specific categories (i.e., proceduralized versus adaptive skill). Results indicated there was less decay for proceduralized skill criteria (δ = –0.23) compared to adaptive skill criteria (δ = –0.43), but their confidence intervals overlapped substantially. In addition, the percent of variance accounted for by

sampling error barely increased after breaking skill criteria into proceduralized and adaptive skill criteria. A relatively small percent of variance explained and overlapping confidence intervals suggests that additional unknown moderators are likely influencing the relationship between decay and different types of skill criteria.

Results regarding Research Questions 2, 3 and 4 together show that criterion type does appear to be an important moderator of decay. Decay for cognitive criteria was stronger than that for skill criteria; declarative and procedural knowledge criteria decayed more than declarative knowledge; and adaptive skill criteria decayed more than proceduralized skill criteria. In addition, findings related to specific types of cognitive criteria were more robust than those for specific types of skill criteria.

Training structure. Research Question 5 asked whether training structure would influence rates of decay. Results showed that the amount of decay increased as the level of structure decreased from high to low ($\delta = -0.18, -0.27$, and -0.68 for high, moderate, and low structure categories, respectively). The greatest difference was found between high and low structure, followed by the difference between moderate and low structure. But the difference between moderate and high structure was less than 0.20. The small percent of variance accounted for by sampling error for low and high structure suggests the presence of additional moderators. Therefore, structure appears to be having a moderating influence, but its influence also depends on other variables.

Practice opportunities. Hypothesis 4 stated that less decay would be associated with opportunities to practice. However, there were not any studies that provided specific information regarding whether trainees had opportunities to practice during the

retention interval. Therefore, Hypothesis 4 could not be tested in the current meta-analysis.

Decay-prevention intervention. Hypothesis 5 stated that post-training decay-prevention interventions would be associated with lower levels of decay. Results presented in Table 7 supported Hypothesis 5; there was no decay when decay-prevention interventions were implemented. In fact, decay-prevention interventions were associated with a boost in performance after the completion of formal training ($\delta = 0.82$). Although the current finding was based on only three available data points that reported information related to decay-prevention intervention. The finding that 100% of the variance was explained by sampling error suggests a low probability of additional moderators that would make a meaningful difference. Thus, it appears that decay-prevention interventions are very worthwhile.

Task-related Factors

Task content. Hypothesis 6 stated that decay would be greater for tasks with high cognitive demands and low psychomotor demands than for tasks with low cognitive demands and high psychomotor demands. As presented in Table 8, results supported Hypothesis 6 indicating more decay for tasks with high cognitive and low physical demands ($\delta = -0.35$) than for tasks with low cognitive and high physical demands ($\delta = 0.23$). 100 percent of the variance was explained by sampling error for tasks with high cognitive and low physical demands, suggesting that the small to moderate amount of decay for such tasks is robust and unlikely to depend on other factors. In contrast, the small positive degree of retention for tasks with low cognitive

and high physical demands is not robust. Only 47 percent of the variance explained by sampling error and the 95% confidence interval ranged from –0.67 to 1.13.

Due to limited data points ($k = 2$) available, interpersonal task demands were excluded from the following moderator analyses concerning task content proposed in Research Questions 6 and 7. Research Question 6 asked whether decay would be related to the strength of each of the content demands. The intent of Research Question 7 was to explore whether decay would differ for tasks that have greater combined demands than for tasks with lower combined demands. To examine both research questions, all three levels of cognitive demands were crossed with all three levels of physical demands to create categories with different combinations of both demands. This is shown in Table 9. Based on the available data, the results regarding Research Question 6 indicated that amount of decay decreased as the physical demands increased across levels of cognitive demands. But no obvious linear trend was found as the cognitive demands increased across levels of physical demands. Tasks with moderate cognitive demands yielded the most decay ($\delta = -0.49$), followed by tasks with high cognitive demands ($\delta = -0.29$), but no decay was found for tasks with low cognitive demands ($\delta = 0.23$). Together, these results show that the degree of both cognitive and physical task demands have an influence on decay. However, the relatively small percent of variance explained by sampling error associated with certain levels of physical (i.e., moderate and high) as well as cognitive demands (i.e., low and moderate) also show that additional moderators may be operating. Therefore, different sets of combined physical and cognitive demands were examined as proposed in Research Question 7.

Regarding Research Question 7, there was no simple trend showing that greater combined demands was associated with either less or more decay compared to fewer combined demands. For instance, for tasks with moderate cognitive demands, there was less decay when there were also physical demands; but for tasks with high cognitive demands, the degree of decay did not differ much depending on the degree of physical demands. The most decay occurred for tasks with moderate cognitive and low physical demands ($\delta = -0.73$); and retention (actually improved performance) was found for tasks with low cognitive and high physical demands ($\delta = 0.23$). Additionally, the effect for tasks with moderate cognitive and low physical demands is fairly robust given the percent of variance explained by sampling variance was 85.22%. It should be noted that findings regarding decay for tasks with high cognitive demands combined with each level of physical demands are also relatively robust because the percent of variance explained by sampling error was 100% for all the effects. Moreover, none of the confidence intervals for the effects for tasks with high cognitive demands overlapped with the confidence interval for the effect for tasks with moderate cognitive demands and low physical demands. Therefore, even though no straightforward pattern was found for the relationship between overall combined task demands and decay, certain combinations of cognitive and physical demands appear to be making a meaningful difference in rates of decay.

Task complexity. Hypothesis 7 stated that decay would be less for more complex tasks than for less complex tasks. As previously described, task complexity in the current meta-analysis consists of four dimensions: closed- versus open-looped, discretion, dynamic complexity, and component complexity. In testing Hypothesis 7,

the pattern of effects for each dimension was examined separately, and then the pattern of effects for combinations of dimensions was examined.

Across the four dimensions separately, the pattern of effect sizes showed mixed support for Hypothesis 7. First, consistent with Hypothesis 7, open-looped tasks ($\delta = -0.30$) were associated with less decay compared to closed-looped tasks ($\delta = -0.44$); however, the difference between the effects was small (0.14). Second, in contradiction to Hypothesis 7, more decay was found for tasks with high discretion ($\delta = -0.45$) compared to tasks with moderate discretion ($\delta = -0.37$), which had more decay compared to tasks with low discretion ($\delta = -0.29$); however, all of the differences between these effects were small (< 0.16). Third, in mixed support of Hypothesis 7, less decay was found for tasks with high dynamic complexity ($\delta = -0.26$) compared to tasks with low dynamic complexity ($\delta = -0.50$) but not compared to tasks with moderate dynamic complexity ($\delta = -0.18$). Fourth, in support of Hypothesis 7, less decay was found for tasks with high component complexity ($\delta = -0.20$); compared to tasks with moderate component complexity ($\delta = -0.37$), which in turn had less decay compared to tasks with low component complexity ($\delta = -0.61$). It is important to note that the effects for all four indicators of high complexity—open-looped, high discretion, high dynamic complexity, and high component complexity—are fairly robust given that the variance explained by sampling error was 100 percent for each of these indicators. None of the effects for the low and moderate complexity levels across the four dimensions was explained much by sampling error ($< 37\%$). These results suggest that decay for complex tasks is less influenced by other moderating factors compared to decay for simpler tasks.

A complete examination of how different combinations of task complexity dimensions are related to decay would involve calculating meta-analytic statistics for all possible combinations of the four dimensions of complexity. However, this was not possible given that no studies (effects) were obtained for many of the combinations. Therefore, discretion, dynamic complexity and component complexity were dichotomized by combining 1 and 2 on the original 0 to 2 rating scale in order to create a $2 \times 2 \times 2 \times 2$ matrix. This matrix is shown in Table 10. Despite collapsing complexity scores in this matter, almost half of the combinations (i.e., combinations involving low discretion) were missing and the combination of open-looped, high discretion, high dynamic complexity, and low to moderate component complexity only included one data point. Given the lack of data points for many combinations of the complexity dimensions, it was not possible to make clear conclusions regarding how decay is related to combinations of complexity and overall task complexity. As shown in Table 10, different combinations of complexity yielding higher combined complexity were not always associated with less decay compared to some combinations yielding lower combined complexity. However, in the majority of comparisons, combinations yielding higher complexity were associated with less decay compared to combinations yielding lower complexity. Also, given high discretion, decay varied more based on combinations of dynamic and component complexity for closed-looped tasks than for open-looped tasks. In sum, no clear support was found for Hypothesis 7, although the general pattern of effects suggest that decay more times than not will be smaller for tasks with higher complexity compared to tasks with lower complexity.

Correlation and Regression Analyses

As an alternative to the traditional meta-analytic approach, correlations between study effect sizes and each of the moderator variables were examined, and a multiple regression analysis was also conducted in order to better examine the unique influence of each moderator variable on decay. As shown in Table 11, study effect size was significantly correlated with criterion type and physical task demands. Skill criteria were associated with less decay than cognitive criteria, and stronger physical demands were associated with less decay. None of the other moderator variables were significantly correlated with study effect size.

There were also statistically significant correlations among many of the moderator variables. First, retention interval covaried with the closed-/open-looped distinction, with open-looped tasks tending to have longer retention intervals than closed-looped tasks. Second, criterion type was significantly correlated with structure, physical task demands, and all task complexity dimensions except for discretion. The covariation between criterion type and physical task demands was particularly noticeable ($r = 0.69$, $p < .01$), indicating tasks with higher physical task demands were more likely to be assessed with skill-based criteria. Third, structure was positively correlated with discretion and dynamic complexity, suggesting training for tasks with higher discretion and dynamic complexity were likely to be more structured. Fourth, cognitive task demands shared significant positive correlations with all four dimensions of task complexity, indicating tasks with higher cognitive demands also tended to be more complex. Fifth, physical task demands were negatively correlated with discretion, but positively correlated with component complexity. In other words, as the physical

demands of a task increased, the number of approaches that can be adopted to accomplish tasks with higher physical demands was limited, but the requirement for number of distinct acts and information cues needed to perform such tasks increased. Last, the closed-/open-looped distinction was significantly correlated with the other three dimensions of task complexity, with open-looped tasks being more complex in other dimensions than closed-looped tasks. Discretion also covaried with component complexity but not dynamic complexity.

Because many of moderator variables were correlated with each other, it could be argued that the meta-analytic results previously reviewed do not provide a clear picture of which variables are predominantly related to decay. Therefore, multiple regression was warranted to examine the unique influence of each moderator variable. As shown in Table 12, the results indicated that the closed-/open-looped distinction (sr^2 = .06), retention interval (sr^2 = .02), and cognitive demands (sr^2 = .02) were the variables that made the strongest contribution to explaining decay. In support of Hypothesis 1, longer retention intervals were associated with more decay ($\beta = -0.17$, p < .05, one-tailed). Stronger cognitive demands were associated with more decay ($\beta = -0.25$, $p < .10$, two-tailed). Although a directional hypothesis was not proposed for the relationship between cognitive demands and decay, this result is partially consistent with Hypothesis 6, which proposed more decay would be found for tasks with high cognitive demands and low physical demands than for tasks with low cognitive demands and high physical demands. In support of Arthur et al.'s hypothesis and Hypothesis 7, open-looped tasks were associated with lower levels of decay ($\beta = 0.37$, p < .01). However, it is important to recognize that interpretations based on these

regression results should be made with caution given the unbalanced distribution of data points across levels of the variables examined.

Discussion

Training is known to be an essential yet expensive endeavor, so organizations investing in training expect durable results or long-term retention of trained knowledge and skill after the completion of training. The impact of poor retention can be particularly salient in the context of training emergency responders or military reserves given that the probability of immediate and frequent application of trained knowledge and skills tends to be fairly low. The only existing quantitative review of decay conducted by Arthur and colleagues (1998) was the first systematic attempt to understand the differential influence of various methodological and task-related factors on decay. Since this initial effort is more than 10 years old and is not without limitations, the current meta-analysis intended to provide an update and extension of the investigation of decay to additional methodological and task-related factors that were not previously examined. In addition, a more sophisticated, multidimensional coding scheme was used in the current meta-analysis to reflect the multi-dimensional nature of task-related characteristics, which also warranted examination of the interaction and covariation among task-related characteristics and methodological factors.

The following discussion will relate findings from the current meta-analysis to those reported in Arthur et al.'s (1998) meta-analysis by comparing the overall effect, as well as moderating effects related to methodological and task-related factors on decay. Then integrative conclusions, theoretical and practical implications will be discussed. In

the end, discussion of limitations of the current meta-analysis will be followed by recommendations for future research.

Overall Effect

Both the current meta-analysis and that conducted by Arthur et al. (1998) indicated decay after the completion of training. However, the overall decay found in the current meta-analysis across all retention interval categories ($\delta = -0.38$; 95% CI = (-1.50, 0.75)) was smaller than that found by Arthur et al. ($\delta = -0.97$; 95% CI = (-2.35, 0.44)). Such differences may partially be due to the lack of overlap in studies included in both meta-analyses. A total of 25 out of all 53 articles included by Arthur et al. (1998) met the current inclusion criteria, out of which 17 did not report relevant statistics needed to calculate effect sizes. Arthur and colleague (1998) adopted the approach recommended by Glass (1981) to convert *p*-values to effect sizes for articles that did not report other relevant statistics (e.g., means, standard deviations, and correlations) needed to calculate more accurate effect sizes (W. Arthur, Jr., personal communication, April 8, 2010). However, it is commonly known that *p*-values are sensitive to study design and sample sizes. More recently, Kraemer (2005) also argued against the use of *p*-values to estimate effect sizes because the exact *p*-values are rarely accurately reported. In addition, different study designs or sample sizes could result in changes in *p*-values but have no effect on the population effect size. Hence, the larger overall effect found by Arthur et al. (1998) might be a result of using *p*-values and sample sizes to calculate effect sizes for some articles that were not included in the current meta-analysis. Nevertheless, it should be noted that effect size from the current meta-analysis was within the range of the confidence interval identified by Arthur et al.

(1998), and vice versa. Therefore, decay indeed occurs, and how to promote knowledge and skill retention is a worthwhile topic for researchers to study, as well as a legitimate concern for practitioners of organizational training.

Methodological Factors

Consistent with the proposed hypothesis and findings from Arthur et al. (1998), results from the current meta-analysis indicated a negative relationship between retention and the length of retention interval (i.e., a positive relationship between decay and length of nonuse). Although effect sizes found for each of the retention intervals were larger in Arthur et al.'s meta-analysis than those in the current investigation, both meta-analyses found minimal or no decay when retention interval was less than 24 hours, and there were no clear linear trends of decay as retention interval extended. Therefore, one can reasonably predict little decay when the length of nonuse is less than a day, but other factors (e.g., task content) may have more influence than the length of nonuse that extends beyond 24 hours.

In line with encoding specificity principle (Tulving, 1983; Tulving & Thompson, 1973) and the identical-elements theory (Thorndike & Woodworth, 1901), the similarity of retrieval condition was identified as the most important moderator by Arthur et al. (1998). But results from the current meta-analysis did not replicate this finding. No support was found for the hypothesis proposing the amount of decay would meaningfully differ depending on whether the conditions between the original and retrieval assessment are similar or different. The discrepancies in findings from both meta-analyses may due to the lack of overlap in studies included in the analyses. Hence,

no conclusive statement can be made regarding the influence of the similarity of retrieval condition on decay.

Compared to the Kirkpatrick taxonomy used by Arthur and colleagues (1998), a more construct-centered taxonomy by Kraiger et al. (1993) was adopted to categorize training evaluation criteria into cognitive and skill-based outcomes in the current meta-analysis. Results from the current meta-analysis showed more decay for cognitive criteria than for skill-based criteria; declarative and procedural knowledge criteria decayed more than declarative knowledge; and adaptive skill criteria decayed more than proceduralized skill criteria.

The greater decay in criteria reflecting a combination of declarative and procedural knowledge compared to criteria only reflecting declarative knowledge may be due to the covariation between criterion type and other task-related factors, such as cognitive, physical demands and closed-/open-loop distinction. Specifically, an inspection of the data points involved in this comparison revealed that when a combination of declarative and procedural knowledge were used to assess learning, 92.9% ($k = 13$) of these studies also involved tasks with moderate cognitive and low physical demands, which decayed the most compared to tasks with other types of combined cognitive and physical demands. In addition, 78.6% ($k = 11$) of these studies were closed-looped tasks that were found to decay more than open-looped tasks. On the other hand, for data points involving only declarative knowledge criteria, 42.9% ($k = 9$) studies involved tasks with moderate cognitive and low physical demands, and 23.8% ($k = 5$) used closed-looped tasks. Therefore, it seems likely that the greater decay for

criterion reflecting a combination of declarative and procedural knowledge was an artifact of task demands.

Nevertheless, although Arthur et al. (1998) followed Kirkpatrick's less construct-focused scheme, their results were similar to the results of the present meta-analysis in that they found more decay for learning criteria ($\delta = -1.07$) than for behavioral criteria ($\delta = -0.78$). Thus, both meta-analyses show that criterion type is an important moderator to consider even though results from both meta-analyses related to criterion type were not directly comparable because different training criterion taxonomies were used.

To address some needs for additional research suggested by Arthur et al. (1998), an additional set of methodological moderators were examined as part of the effort to expand upon their initial efforts. Specifically, Arthur et al. (1998) suggested that future research should focus on appropriate operationalization of initial acquisition because it is the prerequisite of retention. In fact, whether the operationalization of initial acquisition is criterion-based or duration-based was identified as one of the most important moderators based on results from the current meta-analysis, suggesting more decay for criterion-based training than for duration-based training. Like above, one possible explanation for this finding could be differences in task-related factors for studies, in which initial acquisition was criterion-based versus duration-based training. Specifically, given the limited number of studies available, all tasks ($k = 7$) used in criterion-based training were closed-looped tasks with minimal dynamic and moderate component complexity, which were associated with more decay than open-looped tasks or tasks with higher dynamic and component complexity. On the other hand, studies

that were duration-based training had a more even distribution of data points across levels of these task-related factors (e.g., 53.8% were closed-looped and 46.2% were open-looped tasks). Therefore, it seems likely that the less decay for duration-based training than for criterion-based training was an artifact of task complexity.

Training structure is another methodological moderator that was not considered previously by Arthur et al. (1998). Results from the current meta-analysis showed that training with higher structure was found to be associated with less decay. However, at the first glance, such a finding may be viewed as inconsistent with the beneficial effects found in some research regarding learner-centered active learning approaches, which give learners more responsibility for managing their learning with some guidance from the instructional environment. It is important to note that active learning is not the same as discovery learning or open-learning environments, which provide minimal structure and guidance to trainees (Salas et al., 1999; Mayer, 2004). With active learning, some guidance from the instructional environment is provided to ensure that trainees properly focus their attention and fully engage all the training content (Bell & Kozlowski, 2008; Frese et al., 1988). In training using active learning approaches, trainees are offered flexibility and control over their inductive knowledge construction through exploration and experimentation, which do not necessarily translate into effective performance during training, but are likely to facilitate their retention and adaptive transfer (Kozlowski, Chao, & Jensen, 2009; Smith, Ford, & Kozlowski, 1997).

The findings of the present meta-analysis could be viewed as consistent with the active learning literature given that the structure scores from the current meta-analysis had a restricted range from 0 to 9 compared to the possible range of 0 to 12.

Furthermore, only one data point involved a structure score of 9 and no data points involved structure score of 8. Although it is unclear where exactly active learning approaches fall with respect to the low $(0 - 2)$, moderate $(3 - 4)$ and high $(5 - 9)$ structure categories used in the current meta-analysis, it could be argued that the moderate and high structure categories reflected more active learning environments rather than highly structured instructional environments. Nonetheless, results from the present meta-analysis suggest that unstructured training programs (i.e., the low structure category) are not effective in promoting retention. This provides additional support for the possible detrimental effects found for discovery training with no or minimal structure on learning as suggested by Mayer (2004).

One last methodological factor that was not included in Arthur et al. (1998) meta-analysis was the use of post-training decay-prevention interventions because their examination was limited to studies with strict nonuse retention intervals. However, findings related to the beneficial effect of post-training decay-prevention interventions were rather robust because all three studies in the current meta-analysis, which implemented decay-prevention interventions, involved tasks with moderate cognitive and low physical demands, which decayed the most compared to tasks with other types of combined cognitive and physical demands. Therefore, findings from the current meta-analysis provide further support for the value of such interventions in addition to what has been demonstrated by various empirical studies (e.g., Hutchins & Burke, 2006; Richman-Hirch, 2001; Tews & Tracey, 2008).

Task-related Factors

Arthur et al. (1998) showed that decay was greater for cognitive tasks than for physical tasks. The results of the current meta-analysis showed a more complex relationship between decay and cognitive and physical task demands. Rather than dichotomizing tasks into physical versus cognitive tasks, a more sophisticated, multidimensional approach was used to code different levels of cognitive and physical demands, which consequently allowed examinations of different combined cognitive and physical demands to be conducted. Originally, to expand upon Arthur et al. (1998), interpersonal demands were added to the examination of task content beyond cognitive and physical demand, but analyses involving interpersonal demands were not viable in the current meta-analysis due to the limited data points available. Results showed that tasks with high physical and low cognitive demands were retained better than tasks with high cognitive and low physical demands. This finding not only supported the proposed hypothesis, but it was also consistent with the conclusion reached by Arthur and colleagues (1998) regarding decay using the dichotomized distinction between cognitive and physical tasks.

As an extension to Arthur et al.'s (1998) meta-analysis, the multidimensional coding scheme used in the present investigation allowed for an examination as to whether the amount of specific task demands or the extent of combined demands had a differential impact on decay. Current results did not show a simple linear relationship between amount of task demands and decay. No decay was found for tasks with low cognitive but high physical demands. Tasks with moderate cognitive and low physical demands decayed the most. However, minimal differences were found in decay for

tasks with moderate cognitive and moderate physical demands when compared to tasks with moderate cognitive but high physical demands. Furthermore, less decay was found for tasks with high cognitive demands across all levels of physical demands than for tasks with moderate cognitive and minimal physical demands. Such differences are robust because the percent of variance explained by sampling error for all effect sizes were high and their respective confidence intervals did not overlap. Therefore, instead of the amount of combined task demands, it was the combination of cognitive and physical demands that mattered.

Arthur et al. (1998) only focused on the closed- versus open-looped distinction, but did not examine other aspects of task complexity. The current meta-analysis took a multidimensional approach to operationalize task complexity with four dimensions which included the closed- versus open-looped distinction. As a result, examinations of decay as a function of each dimension of task complexity as well as in combination of task complexity dimensions were conducted. Overall, mixed support was found for the hypothesis that there would be less decay for more complex tasks than for less complex tasks. Specifically, less decay was found for open-looped tasks than for closed-looped tasks. This finding was inconsistent with Arthur et al.'s (1998) results, but supportive of their original hypothesis and previous reviews of the literature (e.g., Farr, 1987).

Regarding combined task complexity, although almost half of the combinations could not be examined due to a lack of data points, the general pattern of effects suggests a negative relationship between decay and levels of combined task complexity, with less decay for more complex tasks. Furthermore, given high discretion, decay varied more based on combinations of dynamic and component complexity for closed-

looped tasks than for open-looped tasks. One possible explanation may be due to the nature of closed-/open-looped tasks. A closed-looped task is one with an unambiguous ending that is not contingent on time, and often it involves discrete responses that have a definite beginning and end. On the other hand, an open-looped task does not have a clear and distinct end of the task, and requires constant monitoring of the discrepancies between the current and desired state. Therefore, open-looped tasks are more dynamic and are likely to be higher in component complexity than closed-looped tasks. Overall, the obtained results are consistent with the notion that training involving tasks with greater complexity requires trainees to engage in deeper and more elaborative processing during acquisition, which consequently leads to better retention (Craik & Lockhart, 1972; Schmidt & Bjork, 1992; Wexley & Latham, 2002).

Integrative Conclusions and Implications

In summary, results from the current meta-analysis suggest that decay is a real phenomenon and how to promote the retention of trained knowledge and skill should be on the agenda of both training researchers and practitioners. There is minimal decay when the length of periods of nonuse is less than 24 hours. However, as periods of nonuse extend beyond one day, other factors, especially task-related factors, appear to play more important roles. Although the regression results indicated a slight linear relationship between length of nonuse and decay, conclusions regarding the relationship between length of nonuse and decay should be made with caution. The length of nonuse may not be an important factor to consider in decay without considering task demands.

In particular, upon closer inspection of the data, it appeared that the effects of retention interval covaried with the effects of the combined cognitive and physical

demands (this was not the case with respect to dimensions of task complexity). Accordingly, it was necessary to examine the effect of retention interval on decay separately for each combination of cognitive and physical demands. As shown in Table 13 and Figure 2, there were no clear linear trends of increasing decay as the period of nonuse increased for any of the types of tasks. Tasks with high cognitive and moderate physical demands showed somewhat of a trend of increasing decay across retention intervals, but none of the other combinations showed such trends, and in some cases longer periods of nonuse were associated with less decay or even improved performance. Together, the results shown in Table 13 and Figure 2 indicate that combinations of cognitive and physical demands seem to exert more influence than length of nonuse on decay. In fact, these results suggest that the length of nonuse is not a major determinant of decay compared to the combination of cognitive and physical task demands.

Overall, the pattern of findings suggests that decay is most likely to occur for relatively simple cognitive tasks. In contrast, decay does not appear to be as much of a problem or issue for more cognitively complex tasks or tasks that are predominantly physical in nature. This conclusion is based on the following findings. In the conventional meta-analysis, (a) cognitive criteria were associated with greater decay than skill-based criteria, (b) decay was by far the strongest for tasks with moderate cognitive demands and minimal physical demands than for any other combination of cognitive and physical demands, and (c) higher levels of task complexity were generally associated with less decay. In the regression analysis, decay was (d) stronger with

higher levels of cognitive demands, yet (e) weaker for open-looped versus closed-looped tasks.

Decay is more of a problem for the performance of tasks that mostly require the recall of basic facts or mundane procedures than it is for the performance of tasks that demand information integration or engagement in multiple cognitive processes. In addition, although such a conclusion cannot be generalized to all types of task complexity combinations given the limited data points available, it is important to note that the effects for all four indicators of high complexity—open-looped, high discretion, high dynamic complexity, and high component complexity—are fairly robust, suggesting less decay for more complex tasks. Unlike methodological factors, task-related factors cannot be easily manipulated. Thus, results from the current meta-analysis imply the importance of conducting appropriate task analyses in order to understand different task characteristics, and to consequently incorporate relevant methodological factors that promote the efficiency in training tasks of concern. From researchers' perspective, more research is needed that takes a fine-grained approach to examine task-related factors, as well as possible boundary effects of various training methodologies in relation to different task-related factors. From practitioners' perspective, results from an appropriately conducted task analysis can inform them of when retention is more likely to be a problem, thus if or when specific interventions are needed and likely to be cost-effective in preventing decay.

Limitations and Future Research

There are a few limitations with the current meta-analysis that are important to mention. First, the number of empirical studies included based on the current set of

inclusion criteria was limited when compared to that in Arthur et al. (1998).

Consequently, hypotheses regarding the impact of two important methodological factors

(i.e., degree of overlearning and opportunity to practice) could not be tested. Also due to

the lack of data, hierarchical moderator analyses were not warranted. For example, tasks

that are low in discretion were only reflected in one combination with the other three

task complexity subdimensions. Hence, not all possible combinations of task

complexity subdimensions could be examined in the current meta-analysis.

Furthermore, limited data available did not allow for meaningful interpretation of the

effects of some moderators (e.g., different retrieval condition) on decay. Therefore,

more empirical research is needed in the training literature that examines both

immediate and delayed assessments of learning in relation to potential moderators of

decay. Despite the limited number of primary studies examining decay, another main

cause of small number of studies included in the current meta-analysis was the lack of

necessary statistics to calculate effect sizes. Therefore, researchers should be more

mindful and consistent in reporting necessary statistics in primary studies to warrant

additional meta-analyses with more stable estimates in the future.

Second, additional moderators may still be operating given relatively sizable

standard deviations of δ, along with the small percent of variances accounted for by

sampling error in some of the moderator analyses in the current meta-analysis. For

example, no support was found for the hypothesis suggesting more decay when the

conditions between the original and retrieval assessments were different than when they

were similar. However, no conclusive statement can be made regarding this hypothesis

because the small percent of variance explained by sampling error suggesting additional moderators are likely operating.

Third, the range of structure of training included in the current meta-analysis was rather restricted. Therefore, present findings regarding the high structure category could not be generalized to highly structured (i.e., guided) training. Although results from the current meta-analysis suggest less decay for training with higher structure, future research is needed to explore decay in training programs that fall on the high end of the spectrum of training structure composite in order to examine whether there is a point of diminishing returns of the beneficial effect of structure.

Finally, examination of decay should be extended to the team-level in response to the recent call for studying training and transfer using a multilevel framework (Aguinis & Kraiger, 2009). Although a team comprises of a group of individuals, team performance is not a simple composite of individual performance. Overall team performance is a combination of individual team member's task performance and teamwork performance, such as coordination, team monitoring, and backup behaviors amongst team members (Marks, Mathieu, Zaccaro, 2001; Salas, Burke, & Cannon-Bowers, 2002). Hence, future examination of decay at team level can explore whether decay is more of a function of individual task performance deterioration or degradation of teamwork capabilities.

Conclusion

In sum, there are two primary conclusions that can be drawn from this meta-analysis. One, it appears that task-related factors play a stronger role in decay compared to length of nonuse. That is, decay is not simply a matter of how long individuals go

without performing a task. Two, the small number of studies found shows that substantially more research is needed before a clear theory of skill decay can be articulated. Nevertheless, the results of the present study suggest that task-analysis should be an important part of determining the extent to which decay might indeed be a potential problem and subsequently if decay prevention strategies (both in training [i.e., structure vis-à-vis active learning] and after training) should be incorporated in the design of training programs.

References

* References marked with this symbol studies included in the meta-analysis.

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.

Aguinis, H. & Kraiger, K. (2009). Benefits of training and development for individuals and teams, organizations, and society. *Annual Review of Psychology, 60,* 451-474.

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

Alliger, G. M., Tannenbaum, S. I., Bennett, W., Jr., Traver, H., & Shotland, A. (1997). A meta-analysis of the relations among training criteria. *Personnel Psychology, 50,* 341-358.

Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem solutions. *Psychological Review, 94,* 192-210.

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist, 51,* 355-365.

*Anger, W. K., Patterson, L., Fuchs, M., Will, L. L., & Rohlman, D. S. (2009). *Learning and recall of worker protection standard (WPS) training in vineyard workers*. Unpublished manuscript.

*Anger, W. K., Stupfel, J., Ammerman, T., Tamulinas, A., Bodner, T., & Rohlman, D. S. (2006). The suitability of computer-based training for workers with limited formal education: a case study from the US agricultural sector. *International Journal of Training and Development, 10*, 269-284.

*Anshel, M. H. (1978). Effect of aging on acquisition and short-term retention of a motor skill. *Perceptual and Motor Skills, 47*, 993-994.

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

Arthur, W. Jr., Bennett, W. Jr., & Huffcutt, A. I. (2001). *Conducting meta-analysis using SAS.* Mahwah, NJ: Lawrence Erlbaum Associates.

Arthur, W. Jr., Bennett, W. Jr., Stanush, P. L., & McNelly, T. L. (1998). Factors that influence skill decay and retention: A quantitative review and analysis. *Human Performance, 11,* 234-245.

*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.

* Arthur, W., Jr., Day, E. A., Villado, A. J., Boatman, P. R., Kowollik, V., Bennett, W., Jr., & Bhupatkar, A. (2007). *Decay, transfer, and the reacquisition of a complex skill: An investigating of practice schedules, observational rehearsal, and individual differences* (Rep No. AFRL-RH-AZ-2008-0001). Mesa, AZ: Air Force Research Laboratory.

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

Bassi, L. J. & Van Buren, M. E. (1998). The 1998 ASTD state of the industry report. *Training & Development, 52*, 21-35.

Bell, B. S. & Kozlowski, S. W. J. (2008). Active learning: Effects of core training design elements on self-regulatory processes, learning, and adaptability. *Journal of Applied Psychology, 93,* 296-316.

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

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.) 1999. *How people learn: Brain, mind, experience, and school.* Washington, D.C.: National Academy Press.

Campbell, D. J. (1988). Task complexity: A review and analysis. *Academy of Management Review, 13,* 40-52.

Campbell, J. P. & Kuncel, N. R. (2001), Individual and team training. In N. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran (Eds), *Handbook of Industrial, Work and Organizational Psychology* (pp. 278–313) (London: Sage).

*Campbell, M. (2007). *Error management training from a resource allocation perspective: An investigation of individual differences and the training components that contribute to transfer*. Unpublished doctoral dissertation. Rice University, Texas.

Chen, G., Casper, W. J., & Cortina, J. M. (2001). The roles of self-efficacy and task complexity in the relationships among cognitive ability, conscientiousness, and work-related performance: a meta-analytic examination.

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

Colquitt, J. A., LePine, J. A., & Noe, R. A. (2000). Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology, 85,* 678-707.

Craik, F. I. & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671-684.

Dansereau, D. F., McDonald, B. A., Collins, D. W., Garland, J., Holley, C. D., Diekhoff, G. M., & Evans, S. H. (1979). Evaluation of a teaching strategy program. In H. F. O'Neil, Jr., & C. D. Spielberger (Eds.), *Cognitive and Affective Learning Strategies* (pp. 3-43). New York, NY: Academic Press.

*Davis, F. D. & Yi, M. Y. (2004). Improving computer skill training: Behavior modeling, symbolic mental rehearsal, and the role of knowledge structure. *Journal of Applied Psychology, 89,* 509-523.

*Day, E. A., Arthur, W., Jr., & Gettman, D. (2001). Knowledge structures and the acquisition of a complex skill. *Journal of Applied Psychology, 86,* 1022-1033.

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

DeRouin, R. E., Fritzsche, B. A., & Salas, E. (2004). Optimizing e-learning: Research-based guidelines for learner-controlled training. *Human Resource Management, 43,* 147-162.

Donovan, J. J. & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology, 84,* 795-805.

Dormann, T. & Frese, M. (1994). Error training: Replication and the function of exploratory behavior. *International Journal of Human-Computer Interaction, 6,* 365-372.

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,* 170-177.

*Dupuis, R. j., & Struthers, C. W. (2007). The effects of social motivational training following perceived and actual interpersonal offenses at work. *Journal of Applied Social Psychology, 37,* 426-456.

*Dunn, S. (2000). *Effects of a simulation game on trainees' knowledge and attitudes about age-related changes in learning and work behaviors of older workers.* Unpublished doctoral dissertation. University of North Texas, Texas.

Ebbinghaus, H. (1885). *Memory: A contribution to experimental psychology.* New York, NY: Columbia University.

Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology.* New York, NY: Columbia University.

Eckerman, D. A., Abrahamson, K., Ammerman, T., Fercho, H., Rohlman, D. S., & Anger, W. K. (2004). Computer-based training for food services workers at a hospital. *Journal of Safety Research, 35*, 317-327.

*Eckerman, D. A., Lundeen, C. A., Steele, A., Fercho, H. L., Ammerman, T. A., & Anger, W. K. (2002). Interactive training versus reading to teach respiratory protection. *Journal of Occupational Health Psychology, 7*, 313-323.

*Espejo, J., Day, E. A., & Scott, G. (2005). Performance evaluations, need for cognition, and the acquisition of a complex skill: an attribute-treatment interaction. *Personality and individual differences, 38,* 1867-1877.

*Fante, R. M. (2008). *A comparison of three training methods on the acquisition and retention of automotive product knowledge.* Unpublished Doctoral Dissertation, Western Michigan University, Kalamazoo, MI.

Farina, A. J., Jr., & Wheaton, G. R. (1973). Development of a taxonomy of human performance: The task-characteristics approach to performance prediction. *JSAS Catalog of Selected Documents in Psychology, 3,* 26–27.

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

*Fein, E. C., & Day, E. A. (2004). The PASS theory of intelligence and the acquisition of a complex skill: A criterion-related validation study of Cognitive Assessment System scores. *Personality and Individual Differences, 37*, 1123-1136.

Fleishman, E. A., & Quaintance, M. K. (1984). *Taxonomies of human performance: The description of human tasks.* Orlando, FL: Academic Press.

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

Foxon, M. J. (1993). A process approach to the transfer of training: Part I: The impact of motivation and supervisor support on transfer maintenance. *Australian Journal of Educational Technology, 9,* 130-143.

Foxon, M. J. (1994). A process approach to the transfer of training: Part II: Using action planning to facilitate the transfer of training. *Australian Journal of Educational Technology, 10,* 1-18.

Freebody, P. & Tirre, W. C. (1985). Achievement outcomes of two reading programs: An instance of aptitude-treatment interaction. *British Journal of Educational Psychology, 55,* 53-60.

Frese, M., Albrecht, K., Altmann, A., Lang, J., Papstein, P. V., Peyerl, R. et al. (1988). The effects of an active development of the mental model in the training process: Experimental results in a work processing system. *Behavior and Information Technology, 7,* 295-304.

García, J. A., Moreno, F. J., Reina, R., Menayo, R., & Fuentes, J. P. (2008). Analysis of effects of distribution of practice in learning and retention of a continuous and a discrete skill presented on a computer. *Perceptual and Motor Skills, 107,* 261-272.

Gardner, B., Rose, J., Mason, O., Tyler, P., & Cushway, D. (2005). Cognitive therapy and behavioural coping in the management of work-related stress: An intervention study. *Work & Stress, 19*, 137-152.

Gaudine, A. P. & Saks, A. M. (2004). A longitudinal quasi-experiment on the effects of posttraining transfer interventions. *Human Resource Development Quarterly, 15*, 57-76.

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

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

Goldstein, I. L., & Ford, J. K. (2002). *Training in organizations: Needs assessment, development and evaluation*. (4th ed.). Canada: Wadsworth Group.

*Goodwin, J. E., Grmines, C. R., Eckerson, J. M., & Gordon, P. M. (1998). Effect of different quantities of variables practice on acquisition, retention, and transfer of an applied motor skill. *Perceptual and Motor Skills, 87*, 147-151.

*Grimsley, D. L. (1969). *Acquisition, retention, and retraining: group studies on using low fidelity training devices* (HumPRO Tech. Re. 69-12). Alexandria, VA. Human Resources Resource Organization.

Haerem, T. & Rau, D. (2007). The influence of degree of expertise and objective task complexity on perceived task complexity and performance. *Journal of Applied Psychology, 92*, 1320-1331.

Howell, W. C. & Cooke, N. J. (1989). Training the human information processor: A review of cognitive models. In I. L. Goldstein (Ed.), *Training in Development and Organizations.* San Francisco, CA: Jossey-Bass.

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

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

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: correcting error and bias in research findings.* Thousand Oaks, CA: Sage Publications, Inc.

Hutchins, H. M. & Burke, L. A. (2006). Has relapse prevention received a fair shake? A review and implications for future transfer research. *Human Resource Development Review, 5,* 8-24.

Kirkpatrick, D. L. (1959). Techniques for evaluation training programs. *Journal of the American Society for Training Directors, 13,* 3-9.

Kirkpatrick, D. L. (1976). Evaluation of training. In R. L. Craig (Ed.), *Training and Development Handbook: A Guide to Human Resource Development* (2nd ed.) (pp. 301-319). New York, NY: McGraw-Hill.

Kirkpatrick, D. L. (1996). Invited reaction: Reaction to Holton article. *Human Resource Development Quarterly, 7,* 23-25.

*Konoske, P. J. (1985). *Cognitive factors in learning and retention of procedural tasks.* Unpublished doctoral dissertation. Wayne State University, Detroit, Michigan.

Kozlowski, S. W. J., Chao, G. T., & Jensen, J. M. (2009). Building an inference structure in organizational training: A multilevel approach. In S. W. J. Kozlowski & E. Salas (Eds.) *Learning, training, and development in organizations.* Routledge Academic, NY: New York.

Kraemer, H. C. (2005). A simple effect size indicator for two-group comparisons? A comment on $r_{equivalent}$. *Psychological Methods, 10,* 413-419.

Kraiger, K. (2003). Perspectives on training and development. In W. C. Borman, D. R. Ilgen, R. J. Klimoski, & I. B. Weiner (Eds.), *Handbook of Psychology* (pp. 171-192). Hoboken, NJ: John Wiley & Sons, Inc.

Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology, 78,* 311-328.

*Lim, D. H., & Morris, M. L. (2006). Influence of trainee characteristics, instructional satisfaction, and organizational climate on perceived learning and training transfer. *Human Resource Development Quarterly, 17,* 85-115.

Locke, E.A., & Latham, G.P. (1990). *A theory of goal setting and task performance.* Englewood Cliffs, NJ: Prentice-Hall.

Machin, M. A. (2002). Planning, managing, and optimizing transfer of training. In K. Kraiger (Ed.), *Creating, implementing, and managing effective training and development: State-of-art lessons for practice.* (pp. 263-301). San Francisco, CA: Jossey-Bass.

Mandler, G. & Heineman,S. H. (1956). Effect of overlearning of a verbal response on transfer of training. *Journal of Experimental Psychology, 52,* 39-46.

Mané, A. M., & Donchin, E. (1989). The Learning Strategies Program: An examination of the strategies in skill acquisition. *Acta Psychologica, 71*, 17-22.

Marks, M. A., Mathieu, J. E., Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. *Academy of Management Review, 26*, 356–376.

Marx, R. D. (1982). Relapse prevention for managerial training: a model for maintenance of behavior change. *Academy of Management Review, 7,* 433-441.

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist, 59,* 14-19.

McDaniel, M. A. & Schlager, M. S. (1990). Discovery learning and transfer of problem-solving skills. *Cognition and Instruction, 7,* 129-159.

*McLeod, D. B., & Briggs, J. T. (1980). Interactions of field independence and generatl reasoning with inductive instruction in mathematics. *Journal for Research in Mathematics Education, 11*, 94-103.

*Morin, L., & Latham, G. P. (2000). The effect of mental practice and goal setting as a transfer of training intervention on supervisors' self-efficacy and communication

skills: An exploratory study. *Applied Psychology: An International Review, 49*, 566-578.

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.

Noe, R., Sears, J., & Fullenkamp, A. (1990). Relapse training: Does it influence trainees' posttraining behavior and cognitive strategies? *Journal of Business and Psychology, 4,* 319-328.

*Overdorf, V., Schweighardt, R., Page, S. J., & McGrath, R. E. (2004). Mental and physical practice schedules in acquisition and retention of novel timing skills. *Perceptual and Motor Skills, 99*, 51-62.

Paradise, A. & Patel, L. (2009). *State of the industry: ASTD's annual review of trends in workplace learning and performance*. Alexandria, VA: ASTD.

Pattni, I., Soutar, G. N., & Klobas, J. E. (2007). The impact of a short self-management training intervention in a retail banking environment. *Human Resource Development Quarterly, 18,* 159-178.

*Payne, D. A., Krathwohl, D. R., & Gordon, J. (1967). The effect of sequence on programmed instruction. *American Educational Research Journal, 4*, 125-132.

Quiñones, M. A., Ford, J. K., Sego, D. J., & Smith, E. (1996). The effects of individual and transfer environment characteristics on the opportunity to perform trained tasks. *Training Research Journal, 1,* 29-48.

*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.

*Rawls, J. R., Perry, O., & Timmons, E. O. (1966). A comparative study of conventional instruction and individual programmed instruction in the college classroom. *Journal of Applied Psychology, 50,* 388-391.

*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*, 19-26.

Richman-Hirsh, W. L. (2001). Posttraining interventions to enhance transfer: the moderating effects of work environments. *Human Resource Development Quarterly, 12,* 105-120.

Rivera, R., & Paradise, A. (2006). *State of the industry*. Washington, D. C.: American Society for Training and Development (ASTD).

*Rohlman, D. S., Eckerman, D. A., Ammerman, T. A., Fercho, H. L., Lundeen, C. A., Blomquist, C., & Amger, W. K. (2005). Quizzing and feedback in computer-based and book-based training for workplace safety and health. *Journal of Organizational Behavior Management, 24*, 1-26.

Rohrer, D., Taylor, K., Pashler, H., Wixted, J. T., & Cepeda, N. J. (2005). The effect of overlearning on long-term retention. *Applied Cognitive Psychology, 19,* 361-374.

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

Salas, E., Burke, C. S., & Cannon-Bowers, J. A. (2002). What we know about designing and delivering team training: Tips and guidelines. In K. Kraiger (Ed.), *Creating, implementing, and managing effective training and development* (pp.234-259). San Francisco, CA: Jossey-Bass.

Salas, E., Cannon-Bowers, J. A., Rhodenizer, L., & Bowers, C. A. (1999). Training in organizations: Myths, misconceptions, and mistaken assumptions. *Research in Personnel and Human Resources Management, 17,* 123-161.

Sauer, J., Hockey, G. R. J., & Wastell, D. G. (2000). Effects of training on short- and long-term skill retention in a complex multiple-task environment. *Ergonomics, 43*, 2043-2064.

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

Sharit, J., Czaja, S. J., Hernandez, M., Yang, Y., Perdomo, D., Lewis, J. E., Lee, C. C., & Nair, S. (2004). An evaluation of performance by older persons on a simulated telecommuting task. *Journals of Gerontology: Series B: Psychological Sciences and Social Sciences, 59*, 305-316.

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

Smith, E. M., Ford, J. K., & Kozlowski, S. W. J. (1997). Building adaptive expertise: Implications for training design. In M. A. Quinones & A Dudda (Eds.), *Training for a Rapidly Changing Workplace: Applications of Psychological Research* (pp. 89-118). Washington, D. C.: APA Books.

*Smith-Jentsch, K. A., Campbell, G. E., Milanovich, D. M., & Reynolds, A. M. (2001). Measuring teamwork mental models to support training needs assessment, development, and evaluation: Two empirical studies. *Journal of Organizational Behavior, 22*, 179-194.

Snow, R. E. (1989). Aptitude–treatment interaction as a framework for research on individual differences in learning. In P. L. Ackerman, R. J. Sternberg, & R. G. Glaser (Eds.), *Learning and individual differences* (pp. 13-59). New York: Freedman.

Swanson, R. A. (2001). *Assessing the financial benefits of human resource development.* Cambridge, MA: Perseus.

*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. *Human Factors, 33*, 309–323.

Tews, M. J. & Tracey, J. B. (2008). An empirical examination of posttraining on-the-job supplements for enhancing the effectiveness of interpersonal skills training. *Personnel Psychology, 61,* 375-401.

*Thompson, T. J., Morey, J. C., Smith, S., & Osborne, A. D. (1981). Basic rifle marksmanship skill retention: Implications for retention research (Report no. AD-A134017). U. S. Army Research Institute for the Behavioral and Social Sciences, GA: ARI Field Unit at Fort Benning.

Thorndike, E. L. & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. *Psychological Review, 8,* 247-261.

Trowbridge, M. H. & Cason, H. (1932). An experimental study of Thorndike's theory of learning. *Journal of General Psychology, 7,* 245-260.

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.

*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. & Baldwin, T. T. (1986). Posttraining strategies for facilitating positive transfer: An empirical exploration. *The Academy of Management Journal, 29*, 503-520.

Wexley, K. N. & Latham, G. P. (2002). *Developing and training human resources in organizations.* New York, NY: HarperCollins.

Whitener, E. M. (1990). Confusion of confidence intervals and credibility intervals. *Journal of Applied Psychology, 75,* 315-321.

Wood, R. E. (1986). Task complexity: definition of the construct. *Organizational Behavior and Human Decision Processes, 37,* 60-82.

Wood, R. E., Mento, A. J., & Locke, E. A. (1987). Task complexity as a moderator of goal effects: A meta-analysis. *Journal of Applied Psychology, 72,* 416-425.

*Wood, E., Willoughby, T., Specht, J., Stern-Cavalcante, W., & Child, C. (2002). Developing a computer workshop to facilitate computer skills and minimize anxiety for early childhood educators. *Journal of Educational Psychology, 94,* 164-170.
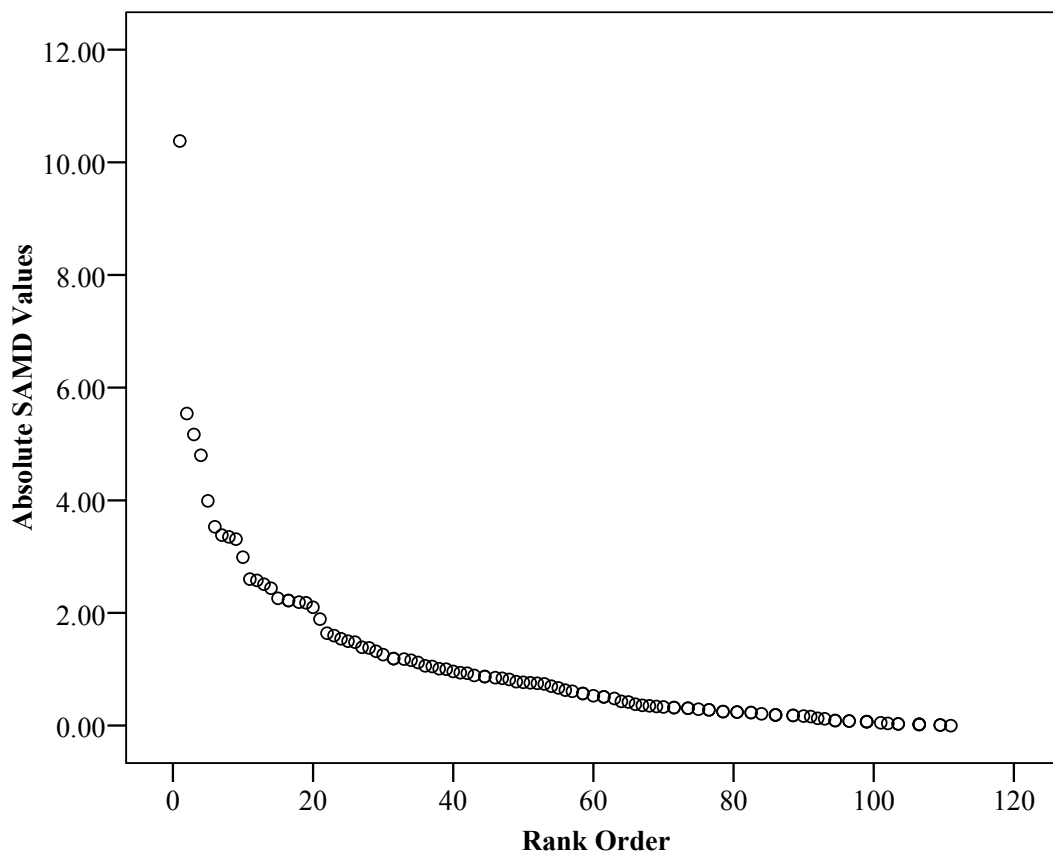
*Figure 1.*Scree Plot for Sample-adjusted Meta-analytic Deviancy Statistics
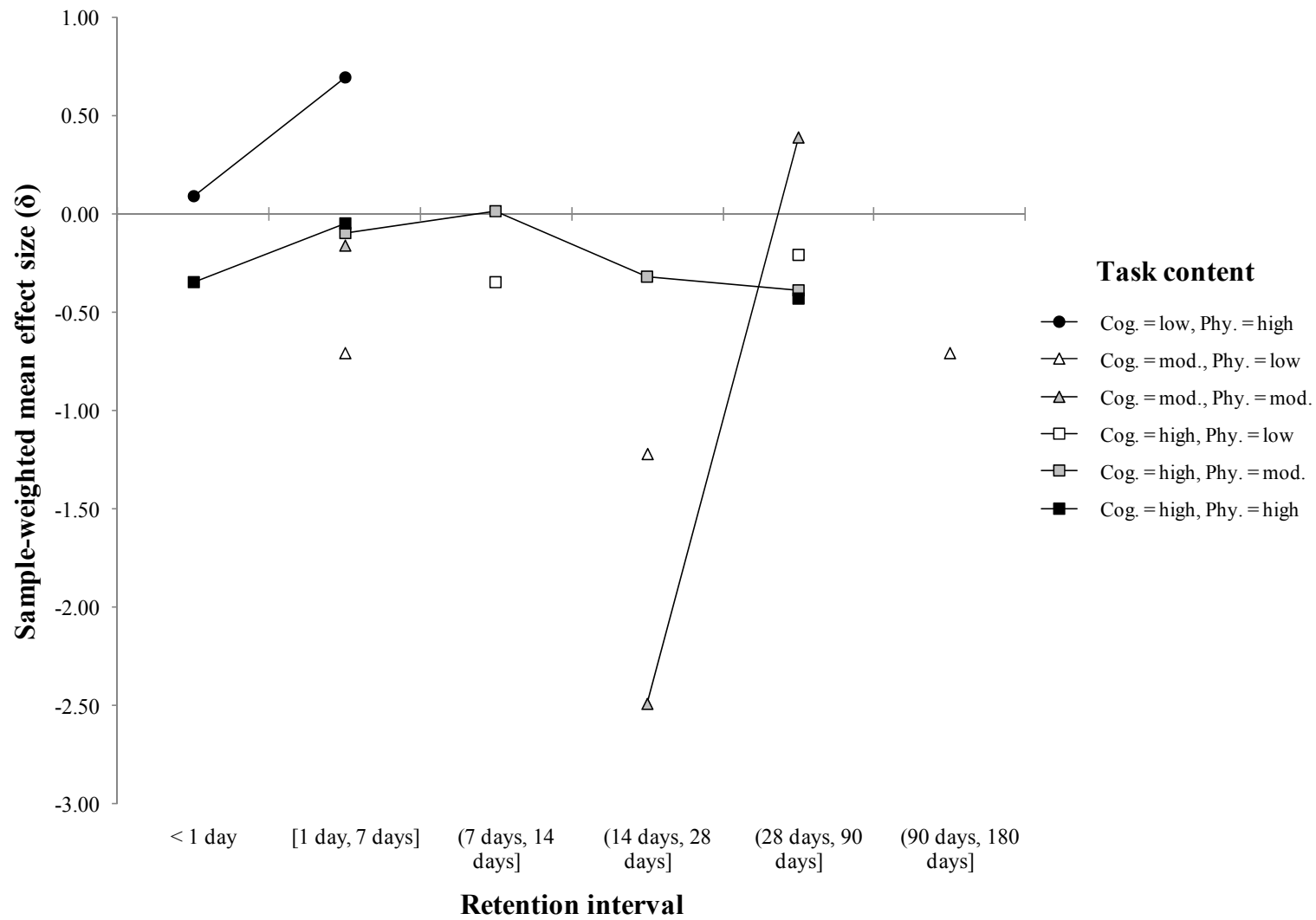
*Figure 2.*Meta-analytic Results for Task Content by Retention Interval

Table 1

Examples of Tasks with Different Content

Task	Cognitive	Physical	Interpersonal
Air traffic control	2	1	1
Chess	2	0	0
Space Fortress	2	2	0
Ring toss	0	1	0
Give a presentation	2	0	1
Word processing	1	2	0
Customer service	1	0	2

Note. 0 = low [minimally applicable/involved]; 1 = moderate [somewhat applicable/involved]; 2 = high [very applicable/involved].

Table 2

Examples of Tasks with Various Levels of Complexity

Task	Open-looped Task	Amount of Discretion	Dynamic Complexity	Component Complexity
Air traffic control	Y	2	2	2
Chess	Y	2	2	1
Space Fortress	Y	2	1	2
Ring toss	N	1	0	0
Give a presentation	N	2	1	1
Word processing	N	1	0	0
Customer service	Y	1	1	1

Note. Y = open-looped; N = closed-looped; 0 = low [minimally applicable/involved]; 1 = moderate [somewhat applicable/involved]; 2 = high [very applicable/involved].

Table 3

Structure Variable Frequencies

Structure Score	Frequency
0	3
1	14
2	21
3	20
4	38
5	5
6	3
7	6
9	1

Note. Scale ranged from 0 to 12.

Table 4

Interrater Agreement for Major Study Variables

Variable	Percent of Agreement
d	100.00
N	100.00
Initial Acquisition	
Overlearning	100.00
Condition of Retrieval	100.00
Type of Evaluation Criteria	
Cognitive	
Declarative Knowledge	98.40
Declarative & Procedural Knowledge	95.27
Skill	
Proceduralized	98.40
Adaptive	98.40
Structure*	92.76
Practice opportunities	100.00
Decay-prevention intervention	100.00
Task Content	
Cognitive	92.60
Physical	92.60
Interpersonal	100.00
Task Complexity	
Open-looped	77.80
Amount of Discretion	100.00
Dynamic Complexity	100.00
Component Complexity	100.00

Note. * = Percent of agreement reported for structure was an average of percent of agreement of all six dimensions. ICC for structure composite score among three coders was .94.

Table 5

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

Note. * = No studies meeting the inclusion criteria were obtained for Interval 7 and 8.

Table 6

Results of Overall Meta-analysis for the Relationship between Retention Interval
and Decay

					% Var.			95% CI	
Retention Interval	*k*	*N*	*δ*	*SD δ*	*Explained*	*Min. d*	*Max. d*	*L*	*U*
Overall	111	3152	−0.38	0.57	30.85	−3.71	1.74	−1.50	0.75
1. < 1 day	10	252	−0.08	0.31	62.97	−0.72	1.74	−0.68	0.53
2. [1 day, 7 days]	48	1283	−0.42	0.27	8.14	−1.62	1.29	−0.94	0.11
3. (7 days, 14 days]	8	259	−0.31	0.00	100.00	−0.40	0.07	−0.31	−0.31
4. (14 days, 28 days]	14	278	−0.84	0.84	24.14	−3.71	0.24	−2.49	0.81
5. (28 days, 90 days]	28	966	−0.24	0.81	15.32	−2.66	0.72	−1.83	1.35
6. (90 days, 180 days]	3	114	−0.71	0.00	100.00	−0.85	−0.60	−0.71	−0.71

Note. k = number of studies. N = number of participants. $δ$ = sample-weighted mean
effect size. SD $δ$ = standard deviation of the estimated true effect size. % Var. explained
= % variance explained by sampling error. Min. d = minimum effect size. Max. d =
maximum effect size. 95% CI = 95% confidence interval (L = lower, U = upper).

Table 7

Meta-Analysis Results for Methodological Moderator Analyses

Moderator Variable	k	N	δ	SD δ	% Var. Explained	Min. d	Max. d	95% CI	
								L	U
Overall	111	3152	−0.38	0.57	30.85	−3.71	1.74	−1.50	0.75
Initial acquisition									
Criterion-based	7	52	−1.43	1.25	33.15	−3.71	0.22	−3.87	1.02
Duration-based	104	3100	−0.35	0.52	34.19	−2.66	1.74	−1.37	0.66
Overlearning [a]	−	−	−	−	−	−	−	−	−
Condition of retrieval									
Similar	108	3004	−0.37	0.58	30.58	−3.71	1.74	−1.52	0.77
Different	3	148	−0.38	0.30	48.22	−0.85	1.11	−0.97	0.21
Type of evaluation criteria									
Cognitive	38	1135	−0.60	0.24	71.05	−1.99	0.47	−1.07	−0.12
Declarative knowledge	21	716	−0.46	0.13	87.42	−1.99	0.17	−0.72	−0.20
Declarative & procedural knowledge	14	359	−0.93	0.00	100.00	−1.62	−0.39	−0.93	−0.93
Skill	73	2017	−0.25	0.66	25.46	−3.71	1.74	−1.54	1.04
Proceduralized	68	1768	−0.23	0.68	25.31	−3.71	1.74	−1.56	1.11
Adaptive	5	249	−0.43	0.42	32.53	−1.01	0.47	−1.24	0.39
Structure									
Low	38	948	−0.68	0.63	30.26	−2.66	0.47	−1.92	0.56
Moderate	58	1443	−0.27	0.23	76.26	−1.99	1.74	−0.72	0.17
High	15	761	−0.18	0.76	12.10	−3.71	0.72	−1.67	1.31

Table 7 Continued

Meta-Analysis Results for Methodological Moderator Analyses

Moderator Variable	k	N	δ	SD δ	% Var. Explained	Min. d	Max. d	95% CI L	U
Practice opportunities [a]	–	–	–	–	–	–	–	–	–
Decay-prevention intervention									
No	111	3152	−0.38	0.57	30.85	−3.71	1.74	−1.50	0.75
Yes	3	376	0.82	0.00	100.00	0.64	0.86	0.82	0.82

Note. k = number of studies. N = number of participants. δ = sample-weighted mean effect size. SD δ = standard deviation of the estimated true effect size. % Var. explained = % variance explained by sampling error. Min. d = minimum effect size. Max. d = maximum effect size. 95% CI = 95% confidence interval (L = lower, U = upper). [a] No studies were found.

Table 8

Meta-Analysis Results for Task-related Moderator Analyses

Moderator Variable	k	N	δ	SD δ	% Var. Explained	Min. d	Max. d	95% CI	
								L	U
Overall	111	3152	−0.38	0.57	30.85	−3.71	1.74	−1.50	0.75
Task content									
Cognitive = low, Physical = mod. [b]	–	–	–	–	–	–	–	–	–
Cognitive = low, Physical = high	8	176	0.23	0.46	46.97	−0.72	1.74	−0.67	1.13
Cognitive = mod., Physical = low	34	880	−0.73	0.17	85.22	−1.99	0.47	−1.07	−0.40
Cognitive = mod., Physical = mod.	21	422	−0.30	0.66	32.46	−3.71	0.47	−1.59	0.98
Cognitive = mod., Physical = high	11	584	−0.25	1.05	6.53	−2.66	0.72	−2.31	1.80
Cognitive = high, Physical = low	9	351	−0.35	0.00	100.00	−1.05	0.11	−0.35	−0.35
Cognitive = high, Physical = mod.	19	301	−0.30	0.00	100.00	−1.68	0.42	−0.30	−0.30
Cognitive = high, Physical = high	7	373	−0.23	0.00	100.00	−0.57	−0.04	−0.23	−0.23
Task complexity									
Open-looped									
Closed-looped	63	1769	−0.44	0.76	20.47	−3.71	1.74	−1.93	1.05
Open-looped	48	1383	−0.30	0.00	100.00	−1.68	1.29	−0.30	−0.30
Discretion [a]									
Low	23	420	−0.29	0.63	36.25	−3.71	0.47	−1.53	0.95
Moderate	66	2256	−0.37	0.62	23.65	−2.66	1.74	−1.60	0.85
High	22	476	−0.45	0.00	100.00	−1.68	0.42	−0.45	−0.45

Table 8 Continued

Meta-Analysis Results for Task-related Moderator Analyses

					% Var.			95% CI	
Moderator Variable	*k*	*N*	*δ*	*SD δ*	*Explained*	*Min. d*	*Max. d*	*L*	*U*
Dynamic complexity [a]									
Low	64	1745	−0.50	0.64	27.27	−3.71	1.74	−1.76	0.75
Moderate	23	833	−0.18	0.58	25.32	−1.55	0.72	−1.31	0.94
High	24	574	−0.26	0.00	100.00	−1.68	0.42	−0.26	−0.26
Component complexity [a]									
Low	30	728	−0.61	0.82	20.64	−2.66	1.74	−2.22	1.01
Moderate	41	1450	−0.37	0.62	23.53	−3.71	0.72	−1.58	0.83
High	40	974	−0.20	0.00	100.00	−1.68	0.42	−0.20	−0.20

Note. k = number of studies. *N* = number of participants. *δ* = sample-weighted mean effect size. *SD δ* = standard deviation of the estimated true effect size. % Var. explained = % variance explained by sampling error. Min. *d* = minimum effect size. Max. *d* = maximum effect size. 95% CI = 95% confidence interval (L = lower, U = upper). [a] Low (0) = minimally applicable/involved; Moderate/Mod. (1) = somewhat applicable/involved; High (2) = very applicable/involved. [b] No studies were found.

Table 9

Meta-Analysis Results for Task Content Moderator Analyses

| | | Physical Demands | | | |
		Low	Moderate	High	Overall
Low	δ (k)	– [a]	–	0.23 (8)	0.23 (8)
	% Var.	–	–	46.97	46.97
	CI	–	–	(−0.67, 1.13)	(−0.67, 1.13)
Moderate	δ (k)	−0.73 (34)	−0.30 (21)	−0.25 (11)	−0.49 (66)
	% Var.	85.22	32.46	6.53	22.74
	CI	(−1.07, −0.40)	(−1.59, 0.98)	(−2.31, 1.80)	(−1.87, 0.90)
High	δ (k)	−0.35 (9)	−0.30 (19)	−0.23 (7)	−0.29 (35)
	% Var.	100.00	100.00	100.00	100.00
	CI	(−0.35, −0.35)	(−0.30, −0.30)	(−0.23, −0.23)	(−0.29, −0.29)
Overall	δ (k)	−0.62 (43)	−0.30 (40)	−0.17 (26)	
	% Var.	77.03	52.47	13.17	
	CI	(−1.04, −0.21)	(−1.19, 0.60)	(−1.71, 1.37)	

Cognitive Demands (row group label)

Note. δ = sample-weighted mean effect size. *k* = number of studies. % Var. = % variance explained by sampling error. *CI* = 95% confidence interval (L = lower, U = upper). Low (0) = minimally applicable/involved; Moderate (1) = somewhat applicable/involved; High (2) = very applicable/involved. [a] No studies were found.

Table 10

Meta-Analysis Results for Task Complexity Moderator Analyses

			Closed-looped		*Open-looped*	
			Amount of Discretion [a]		*Amount of Discretion*	
			Low	**Mod. to Hi.**	**Low**	**Mod. to Hi.**
No Component Complexity — *Dynamic Complexity*	**Low**	δ (k)	– [b]	−0.91 (14)	– [b]	−0.25 (9)
		% Var.	–	13.33	–	60.49
		CI	–	(−2.91, 1.09)	–	(−0.84, 0.35)
	Mod. to Hi.	δ (k)	– [b]	−0.39 (6)	– [b]	0.43 (1)
		% Var.	–	100.00	–	100.00
		CI	–	(−0.39, −0.39)	–	(0.43, 0.43)
			Amount of Discretion		*Amount of Discretion*	
			Low	**Mod. to Hi.**	**Low**	**Mod. to Hi.**
Yes Component Complexity — *Dynamic Complexity*	**Low**	δ (k)	−0.29 (23)	−0.66 (6)	– [b]	−0.42 (12)
		% Var.	63.28	65.57	–	100.00
		CI	(−1.53, 0.95)	(−1.14, −0.17)	–	(−0.42, −0.42)
	Mod. to Hi.	δ (k)	– [b]	−0.18 (14)	– [b]	−0.24 (26)
		% Var.	–	16.48	–	100.00
		CI	–	(−1.46, 1.10)	–	(−0.24, −0.24)

Note. δ = sample-weighted mean effect size. % Var. = % variance explained by sampling error. *CI* = 95% confidence interval. *k* = number of studies. [a] Amount of discretion, dynamic complexity and component complexity were dichotomized by combining moderate and high as mod. to hi, and low as low. Low (0) = minimally applicable/involved; Moderate (1) = somewhat applicable/involved; High (2) = very applicable/involved. [b] No studies were found.

Table 11

Descriptives and Correlations of Effect Sizes and Moderators

	Variable	M	SD	Skewness	Kurtosis	1	2	3	4	5	6	7	8	9
1.	*d*	−0.43	0.74	−1.51	5.87	–								
2.	Retention Interval (weeks)	3.39	4.45	3.08	12.08	− .12	–							
3.	Criterion Type (Cog/Skill) [a]	0.66	0.48	−0.67	−1.58	.23*	−.11	–						
4.	Structure	3.26	1.69	−0.57	0.75	−.01	.19	.19*	–					
5.	Cognitive Task Demands [b]	1.23	0.60	−0.12	−0.44	−.08	.12	−.01	.16	–				
6.	Physical Task Demands [b]	0.79	0.79	0.31	−1.31	.29**	−.04	.69**	.03	−.07	–			
7.	Closed/Open-looped [c]	0.43	0.50	0.28	−1.96	.13	.29**	−.21*	.10	.56**	−.07	–		
8.	Amount of Discretion [b]	0.99	0.64	0.01	−0.50	.01	.11	−.10	.20*	.46**	−.20*	.56**	–	
9.	Dynamic Complexity [b]	0.64	0.82	0.76	−1.08	.04	.06	.31**	.20*	.61**	.13	.45**	.62**	–
10.	Component Complexity [b]	1.09	0.79	−0.16	−1.39	.06	.08	.32**	.12	.59**	.20*	.25**	.02	.50**

Note. * = $p < .05$. ** = $p < .01$. [a] cognitive criterion = 0, skill criterion = 1. [b] 0 = low [not applicable/involved]; 1 = moderate [somewhat applicable/involved]; 2 = high [very applicable/involved]. [c] closed-looped = 0, open-looped = 1.

Table 12

Summary of Multiple Regression Results

Model	B	SE	β	sr^2
Retention Interval (weeks)	−.028	.016	−.167 [†]	.02
Criterion Type (Cog/Skill) [a]	.297	.240	.191	.01
Structure	−.010	.043	−.023	.00
Cognitive Task Demands [b]	−.306	.181	−.247 [†]	.02
Physical Task Demands [b]	.171	.127	.182	.01
Closed-/Open-looped [c]	.546	.197	.366 **	.06
Amount of Discretion [b]	.111	.175	.096	.00
Dynamic Complexity [b]	−.147	.147	−.162	.01
Component Complexity [b]	.101	.129	.109	.00
R^2				.186**

Note. B = Unstandardized regression weights. *SE* = Standard error. *β* = Standardized regression weights. *sr²* = Squared semi-partial correlation. [a] cognitive criterion = 0, skill criterion = 1. [b] 0 = low [not applicable/involved]; 1 = moderate [somewhat applicable/involved]; 2 = high [very applicable/involved]. [c] closed-looped = 0, open-looped = 1. † = $p < .10$. ** = $p < .01$.

Table 13

Meta-Analysis Results for Task Content by Retention Interval

Task Content/Retention Interval	k	N	δ	SD δ	% Var. Explained	Min d	Max d	95% CI L	U
Cognitive = low, Physical = high [a]									
1. < 1 day	6	136	0.09	0.41	52.35	−0.72	1.74	−0.70	0.88
2. [1 day, 7 days]	2	40	0.69	0.29	72.39	0.15	1.29	0.12	1.25
Cognitive = mod., Physical = low									
2. [1 day, 7 days]	27	696	−0.71	0.00	100.00	−1.62	0.29	−0.71	−0.71
4. (14 days, 28 days]	3	60	−1.22	0.43	57.52	−1.99	−0.35	−2.05	−0.38
6. (90 days, 180 days]	3	114	−0.71	0.00	100.00	−0.85	−0.60	−0.71	−0.71
Cognitive = mod., Physical = mod.									
2. [1 day, 7 days]	15	340	−0.16	0.00	100.00	−0.44	0.16	−0.16	−0.16
4. (14 days, 28 days]	3	34	−2.49	0.92	45.71	−3.71	−0.62	−4.29	−0.69
5. (28 days, 90 days]	2	38	0.39	0.00	100.00	0.23	0.47	0.39	0.39
Cognitive = high, Physical = low									
3. (7 days, 14 days]	5	229	−0.35	0.00	100.00	−0.40	−0.35	−0.35	−0.35
5. (28 days, 90 days]	3	102	−0.21	0.16	81.68	−0.74	0.11	−0.53	0.11
Cognitive = high, Physical = mod.									
2. [1 day, 7 days]	2	29	−0.10	0.00	100.00	−0.18	−0.03	−0.10	−0.10
3. (7 days, 14 days]	2	20	0.01	0.00	100.00	−0.09	0.07	0.01	0.01
4. (14 days, 28 days]	7	164	−0.32	0.00	100.00	−0.99	0.24	−0.32	−0.32
5. (28 days, 90 days]	8	88	−0.39	0.00	100.00	−1.68	0.42	−0.39	−0.39

Table 13 Continued

Meta-Analysis Results for Task Content by Retention Interval

					% Var.			95% CI	
Task Content/RI	*k*	*N*	*δ*	*SD δ*	*Explained*	*Min d*	*Max d*	*L*	*U*
Cognitive = high, Physical = high									
1. < 1 day	3	106	−0.35	0.00	100.00	−0.57	−0.13	−0.35	−0.35
2. [1 day, 7 days]	2	178	−0.05	0.00	100.00	−0.07	−0.04	−0.05	−0.05
5. (28 days, 90 days]	2	89	−0.43	0.00	100.00	−0.46	−0.41	−0.43	−0.43

Note. k = number of studies. *N* = number of participants. *δ* = sample-weighted mean effect size. *SD δ* = standard deviation of the estimated true effect size. % Var. explained = % variance explained by sampling error. Min. *d* = minimum effect size. Max. *d* = maximum effect size. 95% CI = 95% confidence interval (L = lower, U = upper). [a] Low (0) = minimally applicable/involved; Moderate/Mod. (1) = somewhat applicable/involved; High (2) = very applicable/involved.