

**A META-ANALYTIC INVESTIGATION OF PROCEDURAL SKILL
RETENTION AND DECAY**

A Dissertation
Presented to
The Academic Faculty

by

Corey E. Tatel

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Psychology

Georgia Institute of Technology
December, 2023

COPYRIGHT © 2023 BY COREY E. TATEL

**A META-ANALYTIC INVESTIGATION OF PROCEDURAL SKILL
RETENTION AND DECAY**

Approved by:

Dr. Phillip L. Ackerman, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Rick P. Thomas
School of Psychology
Georgia Institute of Technology

Dr. Margaret E. Beier
Department of Psychological Sciences
Rice University

Dr. Christopher W. Wiese
School of Psychology
Georgia Institute of Technology

Dr. Ruth Kanfer
School of Psychology
Georgia Institute of Technology

Date Approved: August 22, 2023

ACKNOWLEDGEMENTS

I would first like to thank my advisors, Dr. Phillip Ackerman and Dr. Ruth Kanfer, for their guidance throughout my time in graduate school. I am lucky to have worked with two of the brightest minds in my field who have shaped the way that I think. I would also like to thank Dr. Margaret Beier, Dr. Rick Thomas, and Dr. Christopher Wiese for serving on the committee of this project and offering valuable insights.

I am appreciative of the efforts of the research assistants who contributed to this project: Sarah Burke, Monica Cai, Sidney Demings, Chandler Dennis, Nneoma Eke, Hayden Fulkerson, Zach Goldman, Josephine Joyil, Alan Sherman, Calyn Siver, Giulia Verzino, and Zack Zhang. Without their help in screening over 20,000 abstracts and 3,000 full text sources for eligibility and organizing initial data, this study would not have been possible.

I am forever grateful to my family – Jake, Leslie, and Zak Tatel – for their unwavering support as I’ve pursued my goals and for shaping me into the person that I am today.

I was lucky to share a graduate school experience with a fantastic group of peers who are both great researchers and friends. One of the aspects of my graduate education at Georgia Tech that I enjoyed most was that I not only learned from and shared laughs with my fellow graduate students in the Industrial Organizational Psychology area, but across all areas of psychology. There are too many of you to name – but thank you to all of you for sharing this experience with me. I would especially like to thank Sibley Lyndgaard, who is an incredible partner both in research and in life. I cannot imagine completing this degree without both your emotional and intellectual support.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
LIST OF TABLES	vii
LIST OF FIGURES	viii
SUMMARY	ix
CHAPTER 1. Introduction	1
1.1 Background	1
1.1.1 Procedural Skills	3
1.1.2 Skill Retention in Military and Healthcare Fields	4
1.2 Literature Review	8
1.2.1 Historical Context	8
1.2.2 Critical Questions	13
1.3 A Priori Research Questions	30
1.3.1 Replication	31
1.3.2 Retention Interval	31
1.3.3 Task Moderators	31
1.3.4 Study Design Moderators	32
1.3.5 Initial Acquisition	32
CHAPTER 2. Method	33
2.1 Literature Search	33
2.2 Inclusion Criteria	34
2.2.1 Primary Empirical Studies	34
2.2.2 Sample	34
2.2.3 Task	35
2.2.4 Study Design	35
2.2.5 Performance Data Requirements	35
2.2.6 Language Requirements	35
2.3 Eligibility Screening	35
2.4 Coding Procedure	38
2.4.1 Performance Data	38
2.4.2 Study Design Information	39
2.4.3 Task Information	40
2.4.4 Publication Information	42
2.4.5 Sample Information	42
CHAPTER 3. Effect Size Calculation & Processing	43
CHAPTER 4. Meta-Analytic Approach	48
4.1 Within-Study Aggregation and Independence	48
4.2 RQ 1: Comparison to Arthur et al. (1998) Findings for Procedural Skills	49

4.3	RQ 2: Rate of Decline When Procedural Skills Are Infrequently Used	51
4.4	RQ 3-8: Moderators of the Retention Interval – Effect Size Relationship	52
4.5	RQ 9: Loss of Initial Skill Acquisition Gains	53
4.6	RQ 10: Comparison of Effect Sizes Computed with Different Formulas	53
4.7	Publication Bias	54
CHAPTER 5. Results		57
5.1	RQ 1: Replication of Arthur et al. (1998) Findings for Procedural Skills	57
5.2	RQ 2: Rate of Decline When Procedural Skills Are Infrequently Used	64
5.3	Moderator Analyses	73
5.3.1	Task-Related Moderators	79
5.3.2	Study Design Moderators	86
5.4	RQ 9: Loss of Skill Acquisition Gains	104
5.5	RQ 10: Comparison of Effect Sizes Computed with Different Formulas	109
5.6	Publication Bias	112
CHAPTER 6. Discussion		120
6.1	Research Question 1: Replication of Arthur et al. (1998)	120
6.2	Research Questions 2 & 9: Skill Loss Estimates	122
6.3	Research Questions 3-8: Moderators of Skill Loss Trends	127
6.4	Research Questions 10: Comparison of Different Effect Size Formulas	133
6.5	Limitations and Future Research Directions	134
CHAPTER 7. Conclusions		140
7.1	State of the Skill Retention Literature	140
7.2	Skill Decay Function	140
7.3	Moderators of Skill Decline	141
7.4	Potential Confounds of Skill Retention Estimates	142
APPENDIX A. Search Terms for Each Database		144
APPENDIX B. Task List		146
APPENDIX C. Lab/Artificial vs. Real-World Task Moderator Analysis Models		148
APPENDIX D. All Task Types Moderator Analysis Models		149
APPENDIX E. Open vs. Closed-Looped Task Moderator Analysis Models		151
APPENDIX F. Component Complexity Task Moderator Analysis Models		153
APPENDIX G. Coordinative Complexity Task Moderator Analysis Models		155
APPENDIX H. Dynamic Complexity Task Moderator Analysis Models		157
APPENDIX I. Dexterity Moderator Analysis Models		159
APPENDIX J. Gross Body Coordination Moderator Analysis Models		161

APPENDIX K. Precision/Steadiness Moderator Analysis Models	163
APPENDIX L. Performance Opportunity Moderator Analysis Models	165
APPENDIX M. Task Instructions Moderator Analysis Models	167
REFERENCES	170

LIST OF TABLES

Table 1	– Monte Carlo Simulation Parameters	45
Table 2	– Arthur et al. (1998) Results	60
Table 3	– Replication of Arthur et al. Findings for Accuracy	61
Table 4	– Replication of Arthur et al. Findings for Speed	62
Table 5	– Replication of Arthur et al. Findings for Mixed Performance	63
Table 6	– Accuracy Meta-Regression Model	72
Table 7	– Speed Meta-Regression Model	72
Table 8	– Mixed Meta-Regression Model	73
Table 9	– Predicted Effect Sizes Based on Meta-Regressions	73
Table 10	– Moderator Analyses	77
Table 11	– Narrative Summary of Moderator Analyses	100
Table 12	– Skill Acquisition Meta-Analysis	105
Table 13	– Skill Retention in the Context of Initial Acquisition Gains	108
Table 14	– Meta-Regression of Paired Samples Effect Sizes for Accuracy	111
Table 15	– Meta-Regression of Paired Samples Effect Sizes for Speed	111
Table 16	– Meta-Regression of Paired Samples Effect Sizes for Mixed	112
Table 17	– Moderating Effect of Publication Types on Effect Size Strength	116

LIST OF FIGURES

Figure 1 – Eligibility Screening	37
Figure 2 – Monte Carlo Simulation Results Comparing Effect Size Formulas	47
Figure 3 – Study Retention Interval Frequency	65
Figure 4 – Accuracy Scatterplot	67
Figure 5 – Speed Scatterplot	68
Figure 6 – Mixed Scatter Plot	69
Figure 7 - Laboratory/Artificial vs. Real-World Tasks	90
Figure 8 - All Task Types	91
Figure 9 - Open vs. Closed-Looped Tasks	93
Figure 10 – Gross Body Coordination for Accuracy Performance	94
Figure 11 – Precision/Steadiness for Mixed Performance	95
Figure 12 - Infrequent Performance Opportunities During Retention Interval	96
Figure 13 – Task Instructions	98
Figure 14 - Skill Retention in the Context of Initial Skill Acquisition Gains	107
Figure 15 - Retention Trends for Accuracy Published vs. Grey Literature	115
Figure 16 – Effect Size By Precision Plots	119

SUMMARY

The extent to which procedural skills involving motor components decay over time is an issue that has significant ramifications for the safety and well-being of individuals and society. Prior researchers have concluded that there is a general pattern of skill decay as a function of the length of the retention interval. However, previous researchers relied primarily on studies that leveraged shorter retention intervals than are characteristic of real-world contexts (e.g., days or weeks) and included skills that require both declarative and procedural knowledge. This dissertation presents a new meta-analysis of skill retention that focuses specifically on procedural skills and leverages a recent influx of interdisciplinary literature (e.g., healthcare, sports psychology) consisting of longer retention intervals (e.g., months and years). A broad literature search led to the inclusion of 1,352 effect sizes from 457 sources. Random-effects meta-regression models were computed with retention interval as a predictor of standardized mean differences representing changes in performance between skill acquisition and skill retention for accuracy-based performance measures, speed-based performance measures, and performance measures that were a mix of accuracy and speed. Results indicated that standardized mean differences increased in magnitude by 0.08 per month for accuracy-based performance measures and 0.06 per month for speed-based and mixed performance measures. Initial skill acquisition performance gains were lost between one year and two and half years after they were acquired. Task type, task complexity, infrequent performance opportunities, and task instructions were identified as potentially meaningful moderators of skill decline rates. Findings provide applied audiences with an estimate of how much skill decay can be

expected if skills are not frequently used and therefore, when refresher training should be considered. Important methodological considerations for skill retention research were also identified, including the need to isolate retention performance from relearning effects and the need to account for Speed-Accuracy-Tradeoff functions when interpreting changes in performance over time.

CHAPTER 1. INTRODUCTION

1.1 Background

The retention of skills that are not used consistently is an issue that has significant ramifications for the safety and well-being of society at large. Consider, for example, an adult who originally learned to ride a motorcycle when they were 25 years old, stopped at age 30, and began to ride once again once they reached retirement age. This may be quite dangerous, as data indicate that individuals who take a break from riding show diminished performance in comparison to consistent riders (Symmons & Mulvihill, 2011). If motorcyclists' skill proficiency levels decline without use, there are implications for the safety of the rider, other drivers, and passengers on the road, especially because returning riders are less likely to undertake training than other riders (Haworth et al., 2002). Such data raise the question of whether refresher training should be offered or required for those who have gone extended periods since they were licensed to operate motor vehicles (Ulleberg et al., 2022).

Fortunately, most drivers and motorcyclists likely do not go for extremely long periods without using their skills. When people use their skills frequently, they develop a level of automaticity through exposure and repetition (Fitts & Posner, 1967). This is not always the case for some professions where procedural skills are required, such as healthcare occupations. For some procedures, healthcare providers may go extended periods of time after initial skill acquisition before they are exposed to situations where they must leverage the skills they have been taught (Kelc et al., 2020; Stansfield & Tai, 2021; Stross, 1983). The potential decay of these skills has significant ramifications for

patient safety, especially given that medical facilities that perform procedures more frequently (i.e., with more frequent exposure) tend to provide higher-quality care during those procedures (Brevig et al., 2015). Therefore, it is critical to address questions such as whether refresher training should be provided or recommended for people who do not consistently use skills relevant to their work, how frequently refresher training should be offered, and which skills should be the focus of such refresher training.

Psychological investigations of skill retention date back to the early twentieth century (Hill, et al., 1913; Swift, 1905; 1906) and enough evidence has accumulated to allow for narrative reviews (Adams, 1964, 1987; Farr, 1987) and meta-analyses (Arthur et al., 1998; Wang, 2010/Wang, et al., 2013) that summarize the literature. However, the literature has expanded substantially since a meta-analysis on the topic was last published as a peer-reviewed journal article. The expanding nature of the literature is best illustrated by the results of keyword searches in Google Scholar's database: Using the search term "skill retention" and limiting results to works published prior to the date in which a meta-analysis on skill retention was published as a peer-reviewed journal article (Arthur et al., 1998) yields 1,510 results, while using the same search term and limiting results to works published since 1999 yields 11,800 results¹. There are now nearly eight times as many publications on skill retention as there were 20 years ago, at least within one database. The purpose of this dissertation is to report a new meta-analytic investigation of skill retention that leverages the recent influx of literature on the topic and focuses specifically on procedural skills.

¹ Search results as of July 2023.

1.1.1 Procedural Skills

Psychologists typically classify knowledge as either “declarative” or “procedural” (Ryle, 1949/2000). On one hand, declarative knowledge, which involves “knowing that” competencies (Chi & Ohlsson, 2005), includes factual information and domain knowledge. On the other hand, procedural knowledge, which involves “knowing how” competencies (Polyany, 1976), includes not just knowledge of what things do but also knowledge of how to use them (Willingham & Nissen, 1989).

The critical distinction between declarative and procedural knowledge in the context of skilled performance is that declarative knowledge is necessary, but often not sufficient for developing adequate skill levels (Ackerman, 2008). For example, a novice chef may have memorized the sequential steps involved with preparing a challenging recipe, but that does not necessarily mean that they have the skills necessary to prepare this recipe with precision under significant time pressure (Xu et al., 2020). A professional chef must develop the ‘know-how’ competencies that characterize procedural knowledge (e.g., relative automaticity) in order to complete their job duties.

The focus of the current study is the extent to which procedural skills, rather than declarative knowledge, are retained over time. I define a procedural skill as a learned action that produces a behavioral output involving a non-verbal motor component. Reciting a step-by-step procedure (e.g., a recipe, assembly steps) verbally would not qualify as a measure of procedural skill because the output of the task is solely vocal. Similarly, recognition tests that are often used to assess forms of declarative knowledge (e.g., multiple-choice tests) would not qualify as an adequate assessment of procedural skill.

While individuals possessing relevant procedural skills might be able to accurately describe a step-by-step procedure or recognize a proper behavioral/motor output, neither is evidence of the individuals' ability to produce the required behavioral/motor output themselves. An assessment of a procedural skill must therefore involve the performance of a motor task.

1.1.2 Skill Retention in Military and Healthcare Fields

Two notable fields where resources have been devoted toward understanding procedural skill retention are the military and healthcare. These groups of occupations involve tasks that often require individuals to produce specific behavioral/motor outputs under conditions that require high levels of precision, often under substantial time pressure. For example, an army drone operator must guide a reconnaissance drone to its target destination in a relatively short period of time by adjusting the drone's trajectory based on changing aerodynamic conditions, lighting conditions, enemy positioning, or fuel status (Mairaj et al., 2019). Similarly, an anesthesiologist whose patient is suffering from a collapsed lung must insert an endotracheal tube within the patient's windpipe to keep the airway open and prevent suffocation (Kovacs et al., 2000). In each instance, a lack of precise or untimely production of the required behavioral/motor output could result in severe consequences concerning safety or effectiveness. The medical and military industries have accordingly devoted more attention to issues of procedural skill decay than the broader organizational training literature has.

The military's interest in issues of procedural skill retention has been an enduring one throughout the twentieth and twenty-first centuries. Two of the most comprehensive reviews of the skill decay literature published during the twentieth century found that the

majority of tasks studied that occurred outside of basic laboratory environments were conducted using tasks relevant to those trained by the military (Adams, 1987; Farr, 1987). There are two primary reasons for the military's enduring interest in skill retention over the years. First, the military has a larger volume of personnel to train compared to most employers. The United States Army employed over 800,000 active personnel in 2021 (Duffin, 2022). It has been estimated that it costs the military \$36,000 per recruit for basic training (Niebuhr et al., 2013). That is a substantial financial investment in training and this does not include more advanced training that some personnel complete (e.g., physicians, engineers). Therefore, the military has a vested interest in seeing a return on investment for the large volume of personnel that participates in training.

Second, it is quite common for military personnel to be expected to adequately perform skills for which they were not recently trained. For example, consider the case of reserve army personnel being activated for duty. After initial training, the overall time commitment of reserve personnel is relatively small in most circumstances, meaning that most reserve personnel are not exposed to many opportunities to practice the skills that they acquired during initial training (Wisher et al., 1991). The extent to which these personnel require refresher training prior to activation is an enduring question (Bodilly et al., 1986).

Historically, the military has relied on the assumption that the time, energy, and resources spent to train personnel pays off in the form of higher levels of proficiency and performance (Sabol & Wisher, 2001). However, evidence suggests that skill decay on military tasks is likely when the relevant task is not performed over significant time periods. For example, procedural knowledge required for military tasks (e.g., operating a

gunner position or a missile launcher in a tank) declines with a lack of practice over time, and procedural knowledge is often not adequately redeveloped using simple refresher exercises (Henik et al., 1999). Therefore, the military has an ongoing interest in predicting how rapidly individual procedural skills are forgotten and developing instructional strategies that improve soldiers' retention of what they were originally taught (Wisher et al., 1999).

While traditional research on skill decay was most often conducted in either basic laboratory or military settings, an influx of recent investigations has focused on the extent to which medical professionals retain procedural skills over time. The cases and patients that healthcare providers are exposed to can dictate the extent to which skills obtained during initial training are used or the extent to which they decay (see Orledge et al., 2012). Investigators interested in this issue have leveraged case studies of real-world scenarios in which care providers have gone extended periods without using skills they acquired during their education and training, such as medical residents returning from research appointments (D'Angelo et al., 2018) and military physicians returning to their primary practice (Braun et al., 2014). Each of these case studies suggests the existence of skill decay – residents returning from research appointments and the faculty that supervise them feel that the returning students demonstrate lower levels of proficiency on complex, technical skills such as surgical skills (D'Angelo et al., 2018) and military physicians returning from deployment lack confidence in their ability to conduct acute and routine care procedures, especially when they had limited exposure to the procedures during employment (Braun et al., 2014, $r_{\text{exposure, confidence}} = 0.64$).

Evidence of medical skill decay extends beyond these case studies to more traditional students and care providers. Practicing physicians provide better care when they leverage skills they commonly use (Gillett et al., 2019, $r_{\text{skill proficiency, exposure}} = 0.60$). In contrast, only 5% of medical students were able to pass a first aid and basic life support assessment two years after initial certification (de Ruijter et al., 2014). Interest in skill retention research within the medical field is likely to continue growing due to increasing concerns over the aging physician population and the need to ensure older care providers maintain the knowledge and skills necessary to provide safe and effective care to their patients (Durning et al., 2010; Peisah et al., 2014).

Promising new directions in the area of medical skill decay/retention include building skill decay models that can leverage information such as motor skill level from simulation performance along with characteristics of specific skills in order to predict when and to what extent medical skills are likely to decay (Linde et al., 2018; Linde & Miller, 2019). However, for these approaches to provide utility, there is a need to have an adequate understanding of which type of skills are at the greatest risk of decay. While some literature from the medical domain corroborates prior skill retention research (e.g., low levels of decay over a span of a week, but greater decay over longer time intervals, see Higgins et al., 2021), the literature also contains studies with longer retention intervals than those that characterize much of the previous skill retention literature. For example, many recent studies assess medical students' retention of skills over multiple years of their training (e.g., Boet et al., 2011; Gillett et al., 2019). These studies may go a long way toward increasing the field's understanding of skill retention.

While skill retention has been a long-standing research interest for several areas, the current meta-analysis is designed to further existing research by: (1) Leveraging an influx of new studies, (2) focusing specifically on procedural skills, and (3) including more studies with longer retention intervals than were often available in previous literature.

1.2 Literature Review

In this section, I briefly discuss the most influential work that has been conducted in the area over the past 120 years (since the early 1900s). While this discussion is not exhaustive, it provides context that justifies the purpose of the current study. After providing the historical context, I proceed to discuss five critical questions that inform the current study and the extent to which the existing literature has addressed each of those four questions. My discussion of critical questions draws heavily on findings from the studies and reviews discussed in the historical context portion of the literature review.

1.2.1 Historical Context

1.2.1.1 Early Case Studies

In early studies of skill retention, investigators used themselves or a small number of colleagues as subjects and trained themselves to complete tasks that required motor skills. They tracked their performance as they acquired their initial skills, refrained from using their newly acquired skills for an extended period of time, assessed their performance after a period of non-use, and tracked the amount of time it took for them to achieve similar levels of performance to those that they displayed at the conclusion of initial training.

The tasks used for these studies were relatively simple and allowed for distinct trials across which investigators could track performance. One task that was used across multiple studies was typewriting. Swift (1906) initially trained himself to use a typewriter by practicing over a period of 50 days. He then waited two years and 35 days before retraining himself over a period of 11 days and compared the two learning curves produced by plotting his performance over time. In a later study, investigators acquired typewriting skills over five months of daily practice (Hill et al., 1913). One of the investigators then tested his speed of reacquisition after 25 years without practice (Hill, 1934) and again after 50 years, when the experimenter/subject was over 80 years old (Hill, 1957). Another task utilized by early investigators was a two-ball juggling task, which required continuously tossing one ball in the air while the other remained in the air. Swift (1905) trained two subjects on the juggling task over a span of 42 days of consistent practice and then tested their performance on the task once per month for the next five months, and once more after 18 months. He then followed up with one of the subjects for an additional retention test after six years (Swift, 1910).

The findings of these studies indicated that procedural skills do in fact decay over time. Subjects were typically not capable of achieving the same level of performance that they achieved at the conclusion of their initial acquisition phase in their early attempts after periods of non-use. However, the subjects required less time to relearn the procedural skills than they did to learn them during initial acquisition, which indicated that relearning occurred after extended periods of non-use (notably longer than many experimenters used in coming years). These early studies set the stage for future studies of skill retention.

1.2.1.2 Narrative Reviews

Interest in skill retention grew following early case studies and by the mid-to-late twentieth century, enough evidence had been published to allow for narrative reviews of the existing literature. Jack Adams, who was arguably one of the most notable researchers in the field of motor skill learning, published two relevant reviews separated by 23 years. His seminal review (Adams, 1964) focused primarily on simple motor skills (e.g., reaction time tasks, rotary pursuit tasks) rather than complex skills that are required for many modern occupations. These skills were studied in contexts representative of “basic research and general experimental psychology” (p. 181). His later review was more comprehensive and traced the evolution of theoretical ideas and empirical findings related to how motor skills are learned, remembered, and transferred across three time periods (Adams, 1987). Both reviews led to similar conclusions pertaining to the nature of the skill retention literature – namely, that it was “a domain empty of productive ideas and in which only a little research is done” (Adams, 1987, p. 64-65). One of Adams’ most important discoveries for the purpose of the current study was that while many studies had investigated motor skill acquisition, far fewer had investigated the retention of motor skills over extended periods of time.

Another narrative review was authored by Marshall Farr, whose review was completed around the time of Adams’ more contemporary review (Farr, 1987). Farr’s interest expanded beyond simple motor tasks, as he was interested in “the kinds of knowledge domains and skill that military personnel must master to function effectively on their jobs” (p. S1). Farr (1987) built upon Adams’ work in two primary ways. First, he included a broader variety of skills than Adams did (e.g., cognitive skills in addition to simple motor skills). Second, while the review was qualitative in nature, Farr (1987) used

a numeric rating system that allowed him to communicate estimates of the extent to which particular factors influenced the decay of procedural knowledge skills over time in comparison to other factors revealed by the literature he summarized.

Generally, the three narrative reviews converged on a common description of the literature on skill retention. Both authors agreed that the studies comprising the literature at the time primarily consisted of either relatively simple tasks carried out in laboratory environments or tasks carried out in military settings. Both also outlined two primary theoretical explanations for the existence of knowledge/skill decay. The first was trace decay, which proposed that learners strengthen memory and perceptual traces of proper motor movements during learning and that these traces decay over time without consistent use (see Adams, 1971). The second was interference with retrieval processes, which suggested that skill decay was a result of learning additional skills which interfere with one's ability to properly leverage the original skill acquired (see Adams, 1967). Finally, both authors noted substantial shortcomings of the existing skill-retention literature. Similar to Adams' quote regarding the unproductive nature of skill retention research, Farr (1987) described the literature as "methodologically flawed, lacking a common metric for measuring the degree of learning, the rate of forgetting, and deficient in the theoretical derivation of interpretation" (p. S3).

1.2.1.3 Meta-Analyses

While Adams' (1964, 1987) and Farr's (1987) reviews gave scholars an idea of the general state of the literature, neither provided objective, quantifiable information regarding the trends of skill decay or the relationship between task-related factors and the

extent to which skills are retained. Two meta-analyses conducted in more recent years, however, have done so. The most notable difference between existing meta-analyses and this study is the scope of the skills considered. Whereas the current study considers only procedural skills with a significant non-verbal motor component, past meta-analyses have included skills that lack a significant motor component. For example, tasks such as the memorization of nonsense syllables, letter detection, and visual search were included in previous meta-analyses but not in the current study.

The most well-known quantitative review and meta-analysis related to skill retention/decay is by Arthur et al. (1998). The authors conducted a review of the literature that resulted in 189 data points across 53 articles that included performance data both before and after a period in which participants did not use or practice the skills in question. A more recent meta-analysis of skill retention was conducted, but the authors' scope was narrower than that of Arthur and colleagues because the authors chose to include only studies that investigated organizationally relevant training programs or lab studies designed to be generalized to organizational settings (Wang, 2010/Wang et al., 2013)². Additionally, the authors chose not to convert *p*-values to effect sizes as Arthur et al. (1998) did and were unable to locate 10 of the sources included in the prior meta-analysis. Because of these differences, only eight articles were included in both meta-analyses.

Both meta-analyses gave approximate estimates of the extent to which skills decline over periods of non-use and identified task-related and methodological factors that

² This study was a doctoral dissertation that was eventually published as a chapter. From this point forward, I refer to the findings from this study solely as Wang et al. (2013), although findings are generally present in both the dissertation and the chapter version.

influenced the extent to which procedural skills decay or were retained over time. However, both previous meta-analyses suffer from limitations based on the nature of the literature the respective authors had access to (e.g., shorter retention intervals than are typical of many occupationally relevant skills, insufficient data to test some moderators). Despite these limitations, the trends identified and factors proposed to impact decay rates serve as a starting point for the current study.

1.2.2 Critical Questions

In this portion of the literature review, the findings from the work of historical significance outlined to this point are reviewed in the context of five critical questions that inform the current study, including what previous literature indicates about each of the five questions. The discussion includes a series of factors that have either been demonstrated or proposed to influence skill retention.

1.2.2.1 To What Extent Do Procedural Skills Decay Over Time?

The most fundamental goal of the current study is to quantify the extent to which skills decay over time. The qualitative reviews mentioned in the previous section, which summarize the state of the literature through the mid-1980s, found that performance on tasks that require procedural knowledge is reasonably well retained over time, at least in comparison to tasks that require declarative knowledge (Adams, 1964, 1987; Farr, 1987). However, most studies included in these reviews are characterized by relatively short retention intervals. Of the studies that Adams (1964) reviewed for his initial paper, none used retention intervals of one year or more and most retention intervals could be described in terms of the number of days or weeks rather than in terms of the number of months. For

example, two of the studies that Adams cited to support the conclusion that motor skills are well retained used retention intervals ranging from one to four weeks, which is a much shorter length of time than one would expect some workers to go without leveraging skills that they acquired during their initial education or training.

In his later review, Adams (1987) reviewed studies with similar retention interval lengths. While retention intervals ranged up to two years, most studies that Adams characterized as investigations of “long-term retention” still contained shorter retention intervals than would be expected from occupational groups who must use skills after extended periods of non-use, with the exception of early studies on typewriting and juggling previously mentioned (Swift, 1905, 1906, 1910; Hill 1957; Hill, et al., 1913). For example, two of the longest retention intervals mentioned were 13 weeks (Meyers, 1967) and four months (Mengelkoch et al., 1971). Farr’s (1987) review also corroborated this shortcoming of the skill retention literature, noting that most assessments of long-term skill retention contained modest retention intervals.

From their qualitative reviews, neither Adams (1964, 1987) nor Farr (1987) were able to comprehensively characterize the extent to which the degree of skill retention/decay changes with increasing periods of non-use for two primary reasons: (1) The limited number of studies using longer retention intervals, and (2) the decision not to use quantitative aggregation (e.g., meta-analysis) limited the authors’ ability to characterize decay curves. By leveraging meta-analytic techniques, Arthur and colleagues (1998) were able to overcome the second limitation to provide an estimate of the degree to which procedural skills are retained, or not retained, over time. The Arthur et al. findings indicated that procedural skills do in fact decline over periods of non-use, and the degree to which

they decline increases with longer intervals of non-use. The correlation between the retention interval and the estimated Cohen's D statistic corrected for asymmetric sample sizes (δ) indicating the degree of skill loss was $r = -0.51$.

Immediately following initial skill acquisition (i.e., a retention interval of less than one day), there was little to no skill loss ($\delta = -0.01$). However, after a year or more of non-use, the level of proficiency for procedural skills declined by nearly one and a half standard deviation units ($\delta = -1.40$). The Wang et al. (2013) meta-analysis similarly revealed a pattern of increasing decline over time, although their meta-analytic effect size estimates were more modest than Arthur et al.'s. They found that the degree of skill loss was trivial following initial acquisition ($\delta = -0.08$) and rose substantially with increasing periods of non-use ($\delta = -0.71$ at a retention interval of between 90 and 180 days). The more modest results may have been a result of the difference in scope (i.e., fewer artificial tasks were included because of the focus on only organizationally relevant skills) or because Wang et al. included a smaller number of studies with longer retention intervals.

While these results suggest that skills do decline a substantial amount over time without use, two limitations suggest that the results may not capture the full extent to which skills are subject to decay. First, the two meta-analyses suffered from similar limitations as the work of their predecessors who sought to characterize the same literature – a limited number of existing studies with longer retention intervals. In the Arthur et al. (1998) meta-analysis, 51% of data points had a retention interval of seven days or fewer, while only 8% had a retention interval of longer than 180 days and 2% of data points had a retention interval beyond one year. The Wang et al. (2013) meta-analysis contained *zero* data points with a retention interval greater than 180 days.

Second, the authors of both meta-analyses categorized retention intervals in an asymmetric manner. The retention intervals were classified based on the following categories: Less than a day, between one and seven days, between eight and 14 days, between 15 and 28 days, between 29 and 90 days, between 91 and 180 days, between 181 and 365 days (Arthur et al., 1988 only), and greater than 365 days (Arthur et al., 1998 only). The use of asymmetric retention intervals may have limited the authors' ability to identify trends within the categories they chose. Arthur et al.'s findings suggest that skills decline more rapidly after a period of 90 days than they do between 29 and 90 days after initial acquisition. However, it is also plausible that skills decline to a greater extent in studies that leverage retention intervals between 272 days and 365 days than they do in studies that leverage retention intervals between 180 and 272 days, for example. If this hypothetical example was true, the classification scheme used by previous authors would not have detected such a difference because the results would be collapsed into a larger category. It is also difficult to determine an estimate of skill retention per unit of time because of the asymmetric retention interval categories. The authors likely chose this classification scheme based on the prevalence of studies yielded by their respective literature reviews that used retention intervals of particular lengths and the meta-analytic techniques available at that time. However, the existence of meta-analytic techniques that allow for the treatment of retention interval as a continuous rather than a categorical variable (i.e., meta-regression) as well as a more voluminous literature with longer retention intervals may allow for a more accurate approximation of the extent to which skills decline over time.

1.2.2.2 What is the Relationship Between Speed and Accuracy in Skill Retention?

Performance measures in the skill acquisition and skill retention literature can generally be classified as speed-based or accuracy-based. Speed is usually scored as either completion time or reaction time where higher scores indicate poorer performance or by the number of items or task components completed in a specified period of time. Accuracy is usually scored either on a spectrum from lower levels of accuracy to higher levels of accuracy (e.g., % correct), the number of errors committed, or deviation of performance from an ideal criterion (e.g., Root Mean Square Error).

Past skill retention work has treated whether speed or accuracy was used to characterize performance on different levels of a task-related moderator. Speed vs. accuracy criteria was introduced as a novel contribution by Arthur et al. (1998). The authors hypothesized that performance on speed-based tasks would be better retained than performance on accuracy-based tasks because learners often reach their asymptotic level of accuracy early during skill acquisition, which creates an illusion of mastery. Their results supported this prediction as performance on speed-based tasks showed lower levels of decay compared to performance on accuracy-based tasks (Arthur et al., 1998, $\delta_{\text{Speed}} = -0.33$, $\delta_{\text{Accuracy}} = -1.02$).

However, it is debatable the extent to which speed and accuracy are directly comparable in the context of skill acquisition or retention. The well-replicated Power Law of Practice states that speed increases with practice linearly as a function of the log number of trials (Newell & Rosenbloom, 1981). However, there is no direct analog to the Power Law of Practice for accuracy-based performance. One would likely expect accuracy-based performance to monotonically increase with practice, but there is not sufficient evidence to conclude a ubiquitous pattern of improvement consistent with speed. Additionally, while

tasks can be classified based on whether primary studies scored performance based on speed vs. accuracy, skilled performance in reality is a combination of both speed and accuracy.

Isolating the influences of speed and accuracy performance on performance is problematic because at the individual level of performance, speed and accuracy are negatively related – a phenomenon known as the Speed-Accuracy-Tradeoff-Function (Wicklegren, 1987). The shape of the function can differ, but the general finding is that during skill acquisition, a deliberate focus on accuracy tends to lead to lower completion/reaction times, and a deliberate focus on performing a task as rapidly as possible leads to lower levels of accuracy. The lack of a Power Law of Practice analog for accuracy and the existence of the Speed-Accuracy-Tradeoff-Function combine to suggest that speed and accuracy are not theoretically comparable constructs in the context of skilled performance and may be best examined separately as opposed to as different levels of the same moderator. However, even if they are treated separately, Speed-Accuracy-Tradeoffs make interpreting skill retention results difficult because if a change in performance is observed over time on a task that requires both speed and accuracy, it is difficult to know whether the change in performance can be explained by a change in skill level or by a change in participants' speed-accuracy-tradeoff decisions.

Study design decisions may also influence the nature of how speed-accuracy tradeoffs operate. For example, task instructions may play a critical role in helping participants decide which component to focus on. Consider, for example, a study that plans to report both completion time and number of errors committed for a task in which participants pressed a sequence of four buttons in response to a stimulus. Study designers

may consider instructing participants to enter the sequence as accurately as possible when the stimulus appears, to avoid making mistakes while entering the sequence, or to enter the sequence as quickly and accurately as possible. The choice that the experimenters make may very well determine the strategy that participants engage in, both at acquisition and retention.

1.2.2.3 What Tasks Are Most Prone to Decay?

1.2.2.3.1 Task Types

Task type has most often been classified in skill retention literature as natural tasks (i.e., tasks that would be performed outside of a laboratory) vs. artificial tasks (i.e., laboratory tasks with little generalization to real-world settings). Arthur et al. (1998) classified each of the tasks in their meta-analysis as either natural or artificial. Examples of tasks that were coded as “natural” included typewriting and a simulated lunar landing and examples of tasks that were coded as “artificial” included tracking tasks and mazes³. The authors hypothesized higher levels of retention for natural tasks because natural tasks tend to be more complex than artificial tasks are, and therefore involve higher levels of attention and effort. Additionally, they conjectured that learners would have higher levels of motivation to practice natural tasks frequently and achieve higher levels of performance than they would for artificial tasks, because the learners would perceive relevance for real-world performance. The Arthur et al. findings somewhat supported their predictions, as evidence indicated slightly higher levels of decay for artificial tasks than for natural tasks

³ While a lunar landing task may not be “natural” to some participants, the authors classified the task as natural because it could be natural to some professions (i.e., astronaut).

(Arthur et al., 1998, $\delta_{\text{artificial}} = -0.98$, $\delta_{\text{natural}} = -0.93$). However, the small magnitude of differences between effect sizes limits one from drawing the definitive conclusion that performance on artificial tasks shows meaningfully higher levels of decay. While the small magnitude of differences is surprising, it is noteworthy that 111 of the 178 effect sizes were from artificial tasks. A focus on procedural skills may lead to a larger number of natural tasks that may change the observed difference.

Classifying task types based on natural vs. artificial is a useful strategy because it can aid in interpreting the extent to which findings will generalize to issues of skill retention in the real world. It was especially effective during the early stages of the literature because, as previously mentioned, the majority of early skill retention studies involved either laboratory or military tasks. However, as previously noted, the current literature contains a broader variety of skills than were available at the time of the Arthur et al. meta-analysis, especially when it comes to procedural skills. It may therefore be useful to broaden the “natural” task category to include classifications of skills represented by different occupations in the real world (e.g., medical/dental, sports, military, transportation).

1.2.2.3.2 Task Ability Demands

Previous meta-analyses have primarily defined the content of task demands in terms of physical and cognitive demands. Physical tasks require “muscular strength, exertion of forces, endurance, and coordination,” whereas cognitive tasks require higher-order mental processes such as “perceptual input, mental operations, problem-solving, and decision making” (Arthur et al., 1998, p. 61). Evidence from both meta-analyses suggests that performance on physical tasks is generally better retained than performance on cognitive

tasks (Arthur et al., 1998, $\delta_{\text{cognitive}} = -1.15$, $\delta_{\text{physical}} = -0.75$; Wang et al., 2013, $\delta_{\text{high cognitive, low physical}} = -0.35$, $\delta_{\text{low cognitive, high physical}} = 0.23$). However, tasks that require solely cognitive demands or low physical demands would not meet the current study's definition of a procedural skill because these tasks cannot be proceduralized and/or do not involve significant non-verbal motor components. For example, cognitive tasks such as the memorization of nonsense syllables (included in the Arthur et al. meta-analysis) or the delivery of a presentation (included in the Wang et al. meta-analysis) would not be included in the current meta-analysis. With the current focus on procedural skills with a significant motor component, the critical task content to consider is the nature of the physical and psychomotor demands required, rather than the spectrum of physical vs. cognitive demands required.

Tasks that prior meta-analyses would have classified as physical require psychomotor abilities (abilities that allow for the manipulation or control of objects) or physical abilities (such as balance, flexibility, strength, and endurance). Classic work on the structure of psychomotor abilities has suggested 12 primary factors: Tapping, finger dexterity, rate of arm movement, aiming, arm-hand steadiness, reaction time, manual dexterity, psychomotor speed, psychomotor coordination, discrimination reaction time, postural discrimination, and hand-precision aiming (Fleishman, 1954). Occupational databases have used derivations of these factors to rank ability requirements associated with different jobs. O*Net (National Center for O*NET Development), for example, classifies occupations based on the extent they require psychomotor demands (e.g., arm-hand steadiness, finger dexterity, manual dexterity, wrist-finger speed, control precision, speed of limb movement) and physical demands (e.g., gross-body coordination, gross-body

equilibrium, stamina, strength). These abilities may capture the demands of procedural skills better than traditional classifications of cognitive vs. physical.

1.2.2.3.3 Open-Looped vs. Closed-Looped Tasks

Early literature on the retention of motor skills and procedural knowledge and skills converged on the idea that the degree to which a task is continuous influences the degree of decay (Adams, 1964; Farr, 1987). In fact, one of the primary conclusions from Adams' seminal review of the motor skill literature was that continuous motor skills were retained to a greater extent than discrete motor skills were (Adams, 1964). The distinction between skills that are continuous vs. discrete is often discussed as performance on open-looped vs. closed-looped tasks. On one hand, closed-looped tasks, or discrete tasks, are tasks that involve a clear beginning and end and involve fixed sequences that rarely require adjustment based on changing conditions (Adams, 1971). In closed-looped tasks, skills are leveraged based on a defined sequence of actions until the clearly defined desired outcome is reached, at which point task engagement is discontinued. Stripping or disassembling a rifle is an example of a closed-looped task (Sabol & Wisher, 2001). On the other hand, open-looped tasks, or continuous tasks, are more fluid in nature. There is often not a definitive beginning or end point, and they often require adjustments to the sequential actions required if conditions are altered (Poulton, 1957). A tracking task, which involves maintaining aim on a target throughout fluctuating conditions (e.g., target movement) using an apparatus such as a control stick, is an example of an open-looped task. (Fleishman & Parker, 1962; Gibb & Dolgin, 1989).

Most accumulated evidence leading up to the publication of the Arthur et al. (1998) meta-analysis indicated greater levels of retention for skills involved in open-looped tasks than for closed-looped tasks (Adams, 1964, 1987; Farr, 1987). However, the authors' findings unexpectedly indicated the opposite. They found that performance on open-looped tasks decayed to a greater extent than performance on closed-looped tasks did (Arthur et al., 1998, $\delta_{\text{closed}} = -0.71$, $\delta_{\text{open}} = -1.04$). Given the well-replicated evidence in favor of the retention of open-looped tasks over closed-looped tasks, the authors attributed this finding to a confounding variable: The vast majority of open-looped tasks revealed by their literature review were also classified as cognitive tasks, which were not as well retained as physical tasks. The findings of the more recent meta-analysis were more aligned with previous literature, as they indicated that open-looped tasks were better retained than closed-looped tasks (Wang et al., 2013, $\delta_{\text{closed}} = -0.44$, $\delta_{\text{open}} = -0.30$), which is consistent with the Arthur et al. speculation that their finding was somewhat of an anomaly related to the nature of the tasks in their dataset.

Performance on open-looped tasks may be better retained than performance on closed-looped tasks because open-looped tasks generally allow for more exposure to the repetition of required skills within trials. For example, a tracking task may involve more frequent motor movements to stay on target than many closed-looped tasks require. Schmidt (1975) proposed that motor skill learning is the result of "schemas" which are understood as the relationship between environmental conditions, response specifications of prior actions, sensory consequences of prior actions, and overall outcomes of prior actions. Each time an individual practices a procedural movement, the perceived relationship between each of those four components (or the schema) is strengthened, which

leads to more entrenched memory for procedural skills. Open-looped tasks also involve more initial time investment to learn the skills required, and therefore greater repetition (Naylor & Briggs, 1961). Such tasks also offer more opportunities to engage in procedural movements, which may strengthen learners' schemas and enhance retention (Posner & Keele, 1970). This does raise the question, however, of whether open-looped task performance is truly better retained than performance on closed-looped tasks or whether it is just more susceptible to relearning effects at the time of retention through increased repetition within trials.

1.2.2.3.4 Task Complexity

Another characteristic that distinguishes tasks from one another and may influence the extent to which performance on the task remains stable over time is task complexity. Task complexity was not directly investigated by Arthur et al. (1998). However, it was mentioned in previous qualitative reviews and included in the Wang et al. (2013) meta-analysis. In his review of the literature on long-term retention of knowledge and skills, Farr discussed what he referred to as “the organizational complexity of a task” (Farr, 1987, p. 68), which he defined as the degree of cohesiveness. According to Farr, tasks with higher organizational complexity are most likely to be retained over extended periods of time because such tasks require more intense processing to acquire the skill which in turn increases the rate of overlearning and the stability of the knowledge or skill.

Wang et al. (2013) addressed a limitation of prior literature articulated by Farr (1987), who commented that “there is no satisfactory way of operationally defining” complexity (p. 69), by leveraging a multidimensional definition of task complexity

published around the same time as Farr's review (Wood, 1986). Wood (1986) proposed that three dimensions combine to determine overall task complexity: Component complexity (i.e., the number of distinct steps or actions that are required), coordinative complexity (i.e., the degree to which actions must be sequenced, completed quickly, and completed frequently, as well the intensity of effort required), and dynamic complexity (i.e., the degree to which task conditions are stable or change). For each task included in their meta-analysis, Wang et al. (2013) coded each dimension of complexity as "low," "moderate," or "high" and found an overall pattern of effects consistent with Farr's conjecture that retention is higher for more complex tasks. Specifically, performance decay decreased as both component ($\delta_{\text{low}} = -0.50$, $\delta_{\text{high}} = -0.26$) and dynamic complexity ($\delta_{\text{low}} = -0.61$, $\delta_{\text{high}} = -0.20$) increased⁴. While this is consistent with Farr's reasoning, it is important to note that both Farr and Wang et al. were interested in skills that require both declarative and procedural knowledge. It is possible that low-complexity skills that require mainly declarative knowledge (e.g., brute memorization) may be more susceptible to rapid decay than low-complexity motor skills that can be proceduralized. It is an open question whether these findings extend to procedural skills⁵.

⁴ There were two exceptions to the overall pattern of increased retention with increased task complexity: (1) Retention decreased as a function of coordinative complexity ($\delta_{\text{low}} = -0.29$, $\delta_{\text{moderate}} = -0.37$, $\delta_{\text{high}} = -0.45$), and (2) Retention was higher for tasks with moderate dynamic complexity ($\delta = -0.37$) than for tasks with high dynamic complexity ($\delta = -0.26$). However, the authors still argued that the overall pattern of results indicates increased retention with increasing levels of task complexity.

⁵ There are three additional variables that are commonly discussed in the literature that were not included in the current study: Recognition vs. recall, learning vs. behavioral evaluation criteria, and overlearning. Recognition vs. recall was not included because this task-related variable is relevant to declarative knowledge, rather than procedural skills. Learning vs. behavioral evaluation criteria was not included because it deals with the evaluation of initial training, rather than retention. Overlearning was not included because the construct presents limitations, which are discussed further in the next section.

1.2.2.4 How Does Skill Acquisition Relate to Skill Retention?

While the primary focus of the current study is skill retention, retention is inherently related to acquisition. Most literature that relates retention to initial learning is focused on “overlearning.” Overlearning is the deliberate overtraining of task performance that occurs when an individual engages with additional trials after they have crossed a threshold of performance that defines proficiency (Driskell et al., 1992). The implication is that although the learner may not significantly improve their performance on subsequent trials, the learner supposedly minimizes the extent to which the skills being developed will decline in the future. A meta-analysis of the relationship between overlearning and knowledge retention (including tasks rooted in declarative and procedural knowledge) found that overlearning had a moderate effect on level of skill retention (Driskell et al., 1992, $d = 0.63$). However, the skill retention meta-analyses discussed to this point did not replicate this effect. Arthur et al. (1998) found little evidence for an effect of overlearning on skill retention, although the vast majority of the studies they analyzed (87%) did not contain necessary information to determine the degree of overlearning. Similarly, Wang et al. (2013) could not report effect sizes for overlearning due to insufficient data.

What some researchers have referred to as the influence of overlearning on retention may actually be the result of achieving the autonomous phase of skill acquisition. One of the hallmarks of skill acquisition involves moving toward fast, effortless, and error-free performance (Brown & Carr, 1989). The Fitts and Posner (1967) framework for skill acquisition captures this progression by characterizing three phases of acquisition: Cognitive, associative, and autonomous phases. As learners progress through each phase, they move from engaging in the effortful processing required to learn the rules and

procedures of the task, to applying the rules and procedures with a low degree of error, to increasing automatization of task skills (Ackerman, 1992). Learners in the autonomous phase have reached a rough asymptote with respect to their accuracy, but they continue moving toward higher levels of automatization. In an ideal circumstance, trainees who develop critical procedural skills would advance to the autonomous phase of skill acquisition and require low levels of effort and attention to successfully carry out the task. They could therefore dedicate their cognitive resources toward parallel tasks. However, the reality is that, with a few exceptions (e.g., the military requires a high degree of repetition to ensure their recruits can assemble and disassemble their weapons under high-pressure circumstances, see Sabol & Wisher, 2001), most people do not reach the autonomous stage of skill acquisition during training. Given that learners do not often develop critical skills to a level of automaticity, overlearning may not account for much variance in retention performance except for on relatively simple tasks.

Rather than focusing on overlearning as a moderator, a more useful approach to integrating skill retention with information related to skill acquisition may involve evaluating skill retention *in the context* of acquisition gains. That is, rather than asking the question of ‘does engaging in overlearning lead to better retention?’, one might ask ‘how long does it take to lose initial acquisition gains if skills are infrequently used?’ or ‘what proportion of initial acquisition gains can be expected to decline over a given interval of non-use?’ These are questions that applied researchers would have a vested interest in because of both the financial investment associated with skill training and the data that highlights the benefits that skill training provides for individuals, organizations, and society at large. Answers to these questions may provide information necessary to structure

the timing and content of refresher training to ensure that the benefits of initial skill acquisition are maintained over time.

1.2.2.5 How Do Experiences After Initial Acquisition Influence Retention?

Another critical question regarding skill retention and decay that informs the current study involves experiences that occur between initial skill acquisition and assessment of skill retention. Much of the existing literature on skill decay assumes that decay occurs when individuals learn particular skills and then are not exposed to scenarios in which they must use those skills for some period of time – often referred to as intervals of “non-use” (e.g., Villado et al., 2013). The Arthur et al. meta-analysis did not investigate practice opportunities as an antecedent, primarily because it focused only on intervals of non-use. The authors excluded any study that did not contain “an identifiable interval of nonuse or nonpractice” (Arthur et al., p. 68). This decision is understandable because any exposure to the task may confound measurements of pure retention.

However, intervals consisting of absolutely zero instances of exposure (i.e., pure retention) may not be reflective of what some occupational incumbents experience. Physicians learn an emergency medical procedure, such as intubation, during residency and may encounter, for example, only two patients requiring such procedures in their first five years following residency (Gillett et al., 2019). Some research participants experience performance opportunities as well. Medical students engaged in a retention study may have had access to simulators (e.g., Castevelli et al., 2009; Sinha, et al., 2008) or engaged in internships or rotations that provided infrequent opportunities for skill use that are not systematically measured (e.g., Diederich et al., 2018; Friederichs et al., 2019; Moazed et

al., 2013; Wayne et al., 2006). Some retention studies contain two retention tests separated by two retention intervals (e.g., Gurung et al., 2020; Jang et al., 2020; Oermann et al., 2014), and the first retention test may serve as exposure to the task that confounds the second retention test as a measure of pure retention. The appropriate way to empirically isolate the effects of pure retention from the performance opportunity in this study design would be to have three between-subject conditions: One group that returns only after the first retention interval, a second group that returns after the second retention interval, and a third group that returns after both retention intervals. However, this is not a common study design, and the question of whether the performance opportunity offered by the first retention test influences performance on the second retention test remains an open question. These examples raise both methodological and practical reasons to investigate the extent to which infrequent performance opportunities influence skill retention.

Wang et al. (2013) sought to account for exposure during the retention interval by including opportunities for practice as a methodological factor that influences decrements in performance over time. While they expected opportunities for practice to be associated with less skill decay, they were unable to test this hypothesis because most of the studies revealed by their literature review were somewhat vague concerning which participants were allowed opportunities for practice or encountered circumstances that drew on relevant skills. Despite the lack of meta-analytic evidence supporting the conclusion that practice diminishes decay, evidence from primary studies does support this conclusion (Gillett et al., 2019; Lohre et al., 2021; Sullivan et al., 2019). Evidence even suggests that mental practice without physically engaging with tasks themselves may enhance the retention of skills (McBride & Rothstein, 1979; Kelc et al., 2020; Ong et al., 2021). While Wang et al.

(2013) were unable to derive meta-analytic estimates for the effect of post-learning practice, the expanded skill retention literature allowed for additional insights into whether infrequent practice opportunities influence the degree to which procedural skills are retained over time.

1.3 A Priori Research Questions

As both Adams and Farr described, the skill retention literature is characterized by a lack of integrated theory or consistent empirical findings, which makes forming deductive predictions (i.e., hypotheses) challenging. Additionally, to the extent that consistent evidence does exist, the degree to which it applies to procedural skills, as opposed to skills that require declarative knowledge, is usually an open question because the vast majority of the research reported to this point has dealt with skill retention broadly and collapsed findings across skills that would qualify as procedural along with skills that would not. The ways in which data have been reported in this literature also presented a challenge to proposing hypotheses. When proposing a hypothesis, it is generally ideal to propose not only an expected direction of a relationship but also an effect size indicating the strength of the relationship expected. Much of the psychological literature on skill retention is dated and precedes modern reporting standards. As a result, many of the methodologically sound skill retention studies that used longer effect sizes presented data only as plots or only as central tendency statistics (i.e., without variation data). This makes it difficult to use prior findings to inform a hypothesized effect size.

As a result, the primary objectives of the current study: To describe the state of an interdisciplinary literature, to quantify the degree to which skills are retained or decline

over time, and to explore moderators of the relationship between retention interval length and skill decline, were largely inductive rather than deductive and the study was guided by open research questions. Nine open research questions are posed below. The research questions are used to structure the Meta-Analytic Approach, Results, and Discussion chapters of this dissertation. I use the research questions as sub-headings in each of these chapters and step through each issue individually to illustrate the analyses used to investigate each research question, the extent to which results answer each research question, and the implications of the results for each research question.

1.3.1 Replication

Research Question 1: Given the current study's focus on procedural skills, what results would Arthur et al. (1998) have found had they used similar inclusion criteria?

1.3.2 Retention Interval

Research Question 2: What is the rate of decline when procedural skills are infrequently used?

1.3.3 Task Moderators

Research Question 3: Does task type moderate the relationship between retention interval length and skill decline?

Research Question 4: Does the open vs. closed-looped nature of tasks moderate the relationship between retention interval length and skill decline?

Research Question 5: Does task complexity moderate the relationship between retention interval length and skill decline?

Research Question 6: Do task ability demands moderate the relationship between retention interval length and skill decline?

1.3.4 Study Design Moderators

Research Question 7: Do infrequent performance opportunities during the retention interval moderate the relationship between retention interval and skill decline?

Research Question 8: Does the extent to which task instructions align with how performance was measured moderate the relationship between retention interval and skill decline?

1.3.5 Initial Acquisition

Research Question 9: How much of initial skill acquisition gains are lost over intervals of infrequent use?

CHAPTER 2. METHOD

2.1 Literature Search

In October of 2022, a broad literature search was conducted across interdisciplinary academic databases (Scopus, Google Scholar, Web of Science, PubMed, PsycINFO) in order to identify primary empirical studies that contain data relevant to procedural skill retention and/or decay published up to October of 2022. To obtain grey literature (i.e., unpublished sources such as dissertations, theses, and technical reports), ProQuest and the National Technical Reports Library (NTRL) of the National Technical Information Service (NTIS) were also searched.

A predefined set of key terms expected to yield relevant results was entered into each database. The terms included combinations of terms related to retention/decay (“retention”, “decay”, “deterioration”, “perishability”, “maintenance”, “degradation”) and to the nature of the skills of interest (“procedural”, “psychomotor”, “perceptual”). The specific search terms used differed slightly for each database based on the volume and relevance of the results generated and the ability of each database to handle nested Boolean logic. “Not” terms were also used for some databases to narrow the volume of results yielded. Appendix A contains a comprehensive list of search terms used for each database.

Finally, after initial searches were conducted, an ancestry search was conducted in order to maximize the number of relevant sources included and ensure that appropriate key search terms were not neglected. The references of five focal publications on the topic of skill retention in psychology (Adams, 1964, 1987; Arthur et al., 1998; Farr, 1987; Wang et

al., 2013) and 11 review articles or reports discovered from the initial search (Bryant & Angel, 2000; Frendrick et al., 1988; Gardlin & Sitterley, 1972; Hagman & Rose, 1983; Hurlock & Montague, 1982; Lionetti, 2012; Lyon, 1977; Pendleton, 1986; Pieters & Zaal, 2019; Rowatt & Shlechter, 1993; Wentzel et al., 1983) were reviewed for relevant sources that were not identified in the initial search. Twenty-three additional sources were identified and included in eligibility screening, but the sources did not indicate the need for further refinement of search terms.

2.2 Inclusion Criteria

All studies included in the meta-analysis were required to meet a set of pre-specified inclusion criteria, as follows:

2.2.1 Primary Empirical Studies

All studies were required to be primary empirical studies. No reviews or analyses of data published previously were included.

2.2.2 Sample

All studies were required to have non-clinical (i.e., no cognitive impairment or other disorders) adult samples (ages 17 and up). If studies included a clinical sample with a control group that fit all other inclusion criteria, only the control group was coded. Age 17 was chosen as the minimum age rather than 18 because undergraduate samples from older studies often contained a small number of 17-year-olds. Animal learning studies were also not included.

2.2.3 *Task*

All eligible studies involved a task that required a procedural skill, defined as a task that required a behavioral output involving a significant non-verbal motor component (i.e., the demonstration of hands-on skill). Tasks that were limited to verbal descriptions of step-by-step procedures or recognition tests (e.g., multiple-choice tests) were excluded.

2.2.4 *Study Design*

Studies were required to include the administration of at least two objective skill measures (i.e., not solely self-report measures) separated by a retention interval of at least 24 hours in which participants did not use the relevant skill or infrequently used the relevant skill. Studies were not included if the participants engaged only in observational learning during acquisition (i.e., they did not practice the skill itself) or if the study involved a treatment that may have influenced performance (e.g., sleep deprivation, alcohol consumption) with no control group.

2.2.5 *Performance Data Requirements*

Eligible studies were required to report continuous dependent variables representing performance both at acquisition and at retention, including both central tendency and variation data needed to compute the effect size statistics.

2.2.6 *Language Requirements*

Studies were required to be published in English.

2.3 Eligibility Screening

Database keyword searches yielded a total of 20,379 unique sources. These sources were screened based on the pre-specified inclusion criteria by me, the project advisor, and a team of undergraduate research assistants. Before screening the entire pool of potential sources, 10 sources that represented the comprehensive set of eligibility criteria were evaluated independently by each individual. Each individual was required to retain all eligible sources before proceeding to screen the larger pool of potentially relevant sources, and the exercise was repeated with an additional set of 10 representative sources until this requirement was satisfied.

Eligibility screening was conducted in two phases. In the first phase, all the abstracts of all 20,379 unique sources were evaluated for information that would prevent the sources from satisfying the study's inclusion criteria. A total of 17,331 sources were excluded during abstract screening, leaving 3,048 remaining sources. In the second phase of eligibility screening, the full text of each remaining source was evaluated based on the inclusion criteria. A total of 2,591 sources were excluded during full-text screening, leaving 457 sources that were included in the meta-analysis⁶. A comprehensive summary of the number of sources excluded for each requirement based on PRISMA guidelines (see Moher et al., 2009) is provided in Figure 1.

⁶ In addition to the studies eliminated based on pre-specified inclusions criteria, six studies involving music tasks were excluded because there were not enough studies with skills to warrant inclusion in the meta-analysis and four studies were excluded due to concerns over data quality.

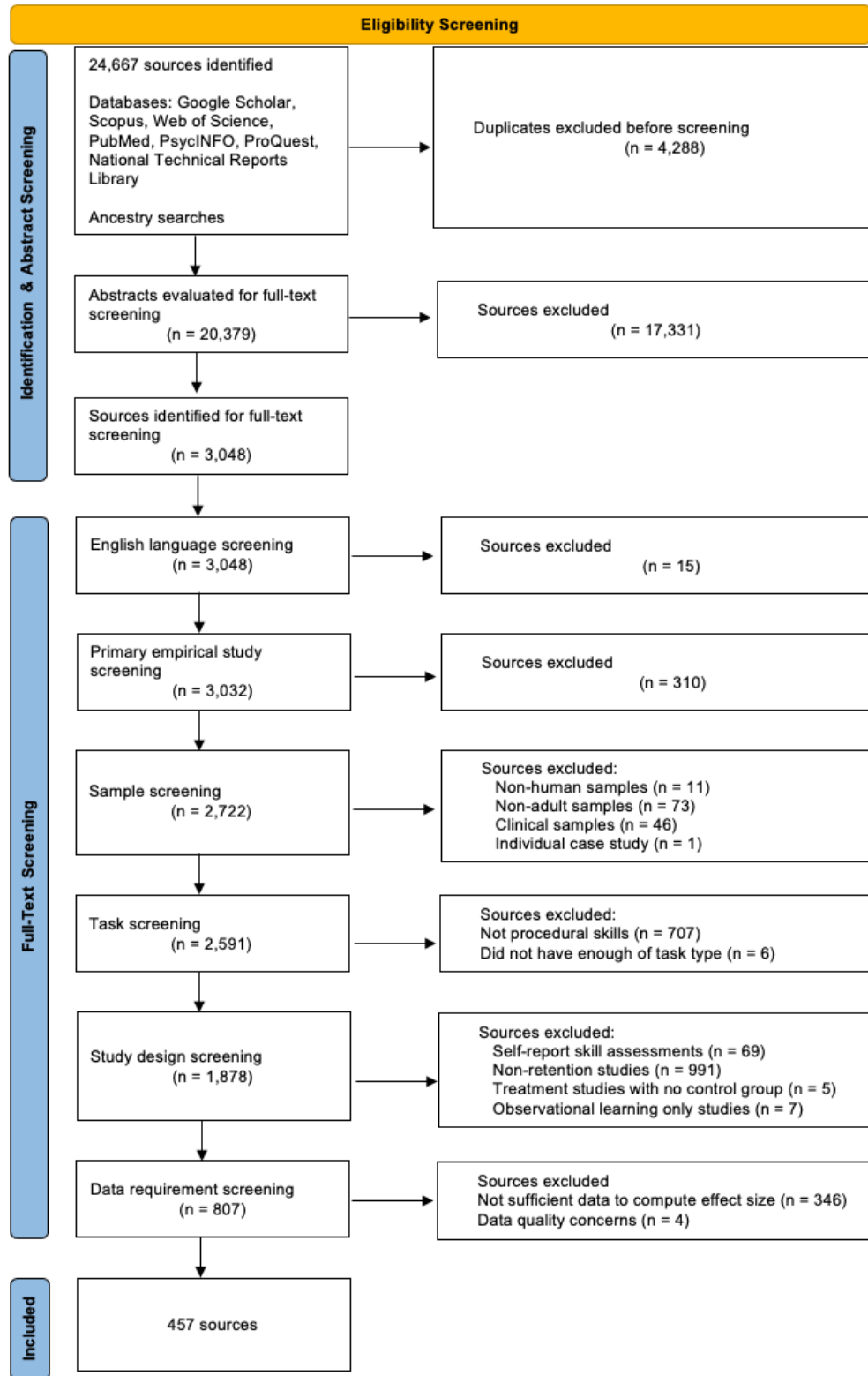


Figure 1 – Eligibility Screening

2.4 Coding Procedure

Following eligibility screening, each of the 457 eligible sources was coded for the information needed to conduct analyses. Each variable that was coded is described below. All variables were coded by a single coder (the author) unless otherwise noted.

2.4.1 Performance Data

Performance at initial acquisition (prior to or in the beginning stages of skill acquisition), final acquisition (at the conclusion of skill acquisition), and retention were recorded. In most cases, means and standard deviations for each time point were recorded. In cases where there was no variance in performance at acquisition (e.g., when all individuals trained until they reached a particular skill level), standard deviation at retention was used as an estimate for underlying standard deviation at acquisition. In some cases, paired-samples Cohen's D statistics were recorded if they were provided or if inferential statistics that could be converted to a Cohen's D statistic (e.g., paired-samples t -value, F -value from a repeated measures ANOVA) were provided. When none of the above were provided, effect sizes were estimated from the information provided whenever possible (e.g., calculation of standard deviation from standard errors of the mean or confidence intervals, estimation of standard deviation by measuring standard error bars on graphs).

When trial-by-trial data were provided, data from the first acquisition trial were coded as initial acquisition performance, data from the final acquisition trial were coded as final acquisition performance, and data from the first retention trial were coded as retention performance. The purpose of coding only the first retention trial whenever possible was to

isolate retention from relearning effects. In cases where trial-by-trial data were not provided, data representing aggregated acquisition or retention trials were coded. All performance data were coded at the group level. For example, if a study reported separate effect sizes for three training groups, three sets of performance data were recorded.

2.4.2 *Study Design Information*

2.4.2.1 Retention Interval

The retention interval(s) associated with the study were coded on a metric of days with the assumption that one month is equivalent to 30 days and 365 days is equivalent to one year. Retention intervals were later converted to a metric of months to aid with interpretation of results.

2.4.2.2 Performance Assessment

Each effect size was coded based on whether the performance measure represented speed, accuracy, or a mixture of both speed and accuracy (“mixed”). When the authors did not provide enough information to determine whether the performance measure was speed-based or accuracy-based, the performance measure was classified as “mixed”.

2.4.2.3 Number of Acquisition Trials

The total number of trials that individuals engaged in during skill acquisition was recorded.

2.4.2.4 Performance Opportunities

Performance opportunity was coded as a binary variable indicating whether or not infrequent performance opportunities were reported or likely during the retention interval. There were two circumstances in which infrequent performance opportunities were coded: (1) When the author mentioned an opportunity for performance (e.g., medical students trained on a surgical task had access to a medical simulator but were not required to use it), and (2) the second of two retention tests administered to the same participants (i.e., the first retention test was an opportunity for performance during the retention interval for the second effect size).

2.4.3 *Task Information*

2.4.3.1 Task Instructions

Task instructions were coded as “speed”, “accuracy”, “both speed and accuracy”, or “not clear”.

2.4.3.2 Task Type

Tasks were classified under one of the following categories: “Laboratory/Artificial”, “Medical/Dental”, “Sports”, “Military”, “Transportation”, or “Miscellaneous” (i.e., not represented by any of the other categories). Military and transportation were later collapsed into a category termed “Military and non-military transportation” because neither category had sufficient data to be included in analyses alone and both categories were deemed to have sufficiently similar skills to justify combining them. A comprehensive list of the tasks classified under each task type is

provided in Appendix B. All task types other than laboratory/artificial tasks were also coded as “Real-world” tasks.

2.4.3.3 Task Characteristics

Task data were further coded under more narrow characteristics. Consistent with prior meta-analyses of skill retention, tasks were coded as either open or closed-looped and were rated as “low” “medium” or “high” on Wood’s (1986) three dimensions of complexity (i.e., component, coordinative, and dynamic). Additionally, tasks were classified based on their ability requirements. Physical and psychomotor abilities requirements associated with jobs listed on Occupational Information Network (National Center for O*Net Development) were consulted and three abilities were chosen for the purposes of this study: Dexterity, gross body coordination, and precision/steadiness. Each ability was coded as a binary variable indicating whether or not the task required the ability.

All task characteristics were completed separately by two independent raters (me and the project advisor). For open vs. closed-looped classifications, disagreements between the raters were resolved through discussion or by reviewing the original publication to gather additional information about the task that was used to make a final determination. For ability requirement classifications, disagreements between the raters were resolved through discussion or by referencing O*Net to determine whether the ability requirement in question was associated with occupations that would typically perform each task. For task complexity dimension rating, classifications were transformed to numeric values where “low” = 1, “medium” = 2, and “high” = 3. In circumstances where ratings differed by only one, the average of the two ratings was taken. For example, if one rating indicated

low component complexity and one rating indicated medium component complexity, a score of 1.5 was used as the rating for component complexity of the task. In circumstances where ratings differed by more than one point (i.e., one rater classified the complexity dimension of the task as low and one rater classified the complexity dimension of the task as high), disagreements were resolved through discussion or reviewing the original publication for more detailed task information.

2.4.4 Publication Information

The year of publication was recorded and the publication type was coded. Publication type was coded as either “Published” (journal articles and book chapters) or “Grey Literature” (theses/dissertations, conference proceedings papers, technical reports).

2.4.5 Sample Information

Sample information coded included sample type, sample age, and sample size. Sample type was classified into one of the following categories: Students, general population or unspecified, medical/dental, military, athletes, or other. Sample age was coded as the mean or median of age of the sample, depending on which was provided. Finally, sample sizes both at the time of skill acquisition and at the time of the skill retention assessment were recorded.

CHAPTER 3. EFFECT SIZE CALCULATION & PROCESSING

The majority of performance data were coded as means and standard deviations for initial acquisition, final acquisition, and retention. The means and standard deviations were used to compute the effect size Cohen's D using the formula generally used to calculate the effect size between independent samples. While it would have been ideal to use the Cohen's D formula for paired samples to account for dependency between skill measurements of the same subjects, $d = t [2(1-r)/n]^{1/2}$ (Dunlap et al., 1996), the formula requires the correlation between measures and a paired t -statistic, and these were rarely provided in the literature. The use of the independent samples Cohen's D formula is consistent with approaches used in prior skill retention meta-analyses (Arthur et al., 1998; Wang et al., 2013). The direction of the effect sizes, as indicated by Equations 1 and 2 below, were structured such that gains in skill were indicated by positive effect sizes and declines in skill were indicated by negative effect sizes.

$$d_{Retention} = \frac{Mean_{Retention} - Mean_{Final\ acquisition}}{Pooled\ Standard\ Deviation} \quad (1)$$

$$d_{Acquisition} = \frac{Mean_{Final\ acquisition} - Mean_{Initial\ acquisition}}{Pooled\ Standard\ Deviation} \quad (2)$$

However, there were other effect sizes in the dataset that were not calculated using means and standard deviations. These data were reported by the authors as either Cohen's D statistics, t -values, or F -values. F -values were converted to t -values by calculating the square root of the F -value and paired t -values were converted to Cohen's D statistics based on the formula $d = t/(n)^{1/2}$ (Lakens, 2013; Rosenthal, 1991). Once again, the ideal

conversion formula would have been the one provided by Dunlap et al. (1996), but the use of this formula was not possible because authors rarely reported correlations between performance at acquisition and performance at retention. In total, 1,281 effect sizes coded were calculated using the independent samples formula, 61 were converted from t -values (or F -values converted to t -values), and 10 were reported as Cohen's D statistics by authors.

In order to investigate the potential consequences of combining effect sizes computed from different formulas, three Monte Carlo simulations were run in which the three Cohen's D formulas in question (the formula for independent samples, the formula for paired samples that takes the correlation between measures into account, and the formula that converts paired t -statistics to Cohen's D) were varied under four of parameters: (1) The correlation between skill acquisition and retention, (2) the mean difference between acquisition and retention, (3) the ratio of standard deviations between acquisition and retention, and (4) the sample size. The specific parameters that were used are provided in Table 1. For each combination of parameters (a total of 128 combinations), a simulation was run 500 times to account for robustness and variability. The output of each simulation was the average Cohen's D statistic and the randomly generated parameters for each of the 500 runs. The results of the three simulations were compared to determine the conditions under which the formulas were likely to yield similar or different results.

Table 1 – Monte Carlo Simulation Parameters

Parameters	Values Used
Correlations between acquisition and retention	0.20, 0.40, 0.60, 0.80
Mean differences between acquisition and retention	0.00, 0.50, 1.50, 3.00
SD ratios for acquisition vs. retention	1.00, 2.00
Sample size	10, 20, 50, 100

Note. This table defines the specific parameters that were varied in each Monte Carlo simulation.

The results of the three simulations across parameter values are plotted in Figure 2. As the figure displays, the formulas yielded quite different effect sizes under some conditions, particularly when the correlation between acquisition and retention was high and when the mean difference in performance between acquisition and retention was high. When the correlation between acquisition and retention was 0.80, the independent samples formula yielded an average effect size of -1.07, the conversion from a paired *t*-value yielded an average effect size of -1.62, and the paired samples formula yielded an average effect size of -2.59. When the mean difference in performance between acquisition and retention was three, the independent samples formula yielded an average effect of -2.56, the conversion from a paired *t*-value yielded an average effect size of -2.82, and the paired samples formula yielded an average effect size of -4.36.

For the paired *t*-values that were converted, the differences at high correlations between measures were large enough that it would not be justifiable to combine them with the effect sizes calculated using the independent samples formula, especially because one would expect the correlation between performance at acquisition and performance at retention to be large. There is ultimately no way to be certain how the Cohen's *D* statistics reported by authors of the primary sources were calculated. Lakens (2013) outlines six

possible formulas for calculating Cohen's D . It is somewhat unlikely that most authors accounted for the correlation between measures in the computations of Cohen's D , particularly if the sources were published prior to the publication of the Dunlap et al. article in 1996. However, if the authors did account for the correlation between acquisition and retention in their computations (which would be a more accurate estimation because it accounts for the dependence between measurements), the differences from the effect sizes computed using the independent samples formula would have been even larger.

Based on the results of these simulations and the fact that correlations between acquisition and retention were not frequently reported in the skill retention literature, a decision was made to treat the effect sizes calculated from the independent samples formula separately from the effect sizes converted from paired samples t -values and Cohen's D values that were reported by authors. While the independent samples formula is admittedly not the ideal formula for a repeated-measures design such as a skill retention study, the study's primary research questions (i.e., Research Questions 1-9) were assessed using the independent samples formula for three primary reasons: (1) Only 5% of the effect sizes coded were not computed using the independent samples formula and the small sample size would likely lead to a less robust estimate of skill decay and lack sufficient power to detect moderator effects, (2) the means and standard deviations from the final acquisition trial and first retention trial used to compute effect sizes using the independent samples formula allowed for better isolation of retention from relearning effects than effect sizes that aggregated across retention trials did, and (3) the use of the independent samples formula is consistent with approaches taken in prior skill retention meta-analyses (Arthur et al., 1998; Wang et al., 2013). However, the results of these simulations pose the question

of how skill retention trends compare between those computed with the independent samples formula and those not computed with the independent samples formula. While this question was not anticipated a priori, it was included as an additional research question based on the simulation results:

Research Question 10: How do skill retention trends and effect sizes compare between effect sizes computed with different formulas?

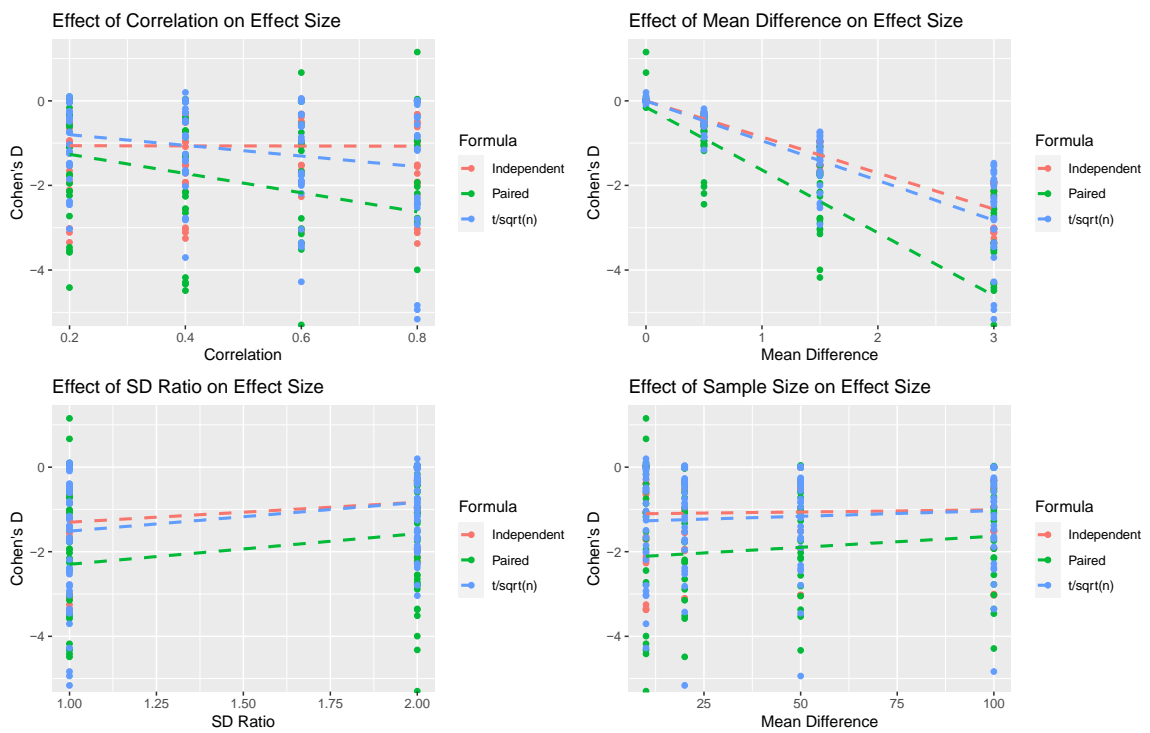


Figure 2 – Monte Carlo Simulation Results Comparing Effect Size Formulas

Note. Results of Monte Carlo simulations that investigated the results of three Cohen's *D* formulas under varying observations of the correlation between measures, mean differences between measures, SD ratio between measures, and sample size.

CHAPTER 4. META-ANALYTIC APPROACH

In this chapter, I describe the meta-analytic approach that I used to test the 10 research questions of the current study. First, I discuss decisions that were made with respect to the aggregation of effect sizes to maximize the independence of observations and provide justification for why some effect sizes from the same sample were not aggregated and were instead analyzed separately or treated independently. Next, I report the meta-analytic models that were computed to answer each of the research questions proposed in the Introduction chapter. Finally, I report a two-pronged approach used to investigate the existence of publication bias in the meta-analytic models tested. The results of these analyses are presented in the next chapter, but the methods and analyses used to obtain results are reported here.

4.1 Within-Study Aggregation and Independence

While an important consideration of meta-analyses is the independence of observations and it is usually advisable to aggregate effect sizes such that all participants are represented by only one effect size (Schmidt & Hunter, 2015), decisions regarding aggregation in the current meta-analysis were complicated by both the lack of theoretical comparability of some effect sizes and study designs that used multiple retention intervals or multiple tasks that assessed different procedural skills.

The current study sought to meta-analyze the strength of effect sizes across time for three types of performance measures: Speed, accuracy, and mixed. Due to the lack of evidence that speed and accuracy are theoretically comparable constructs (e.g., the Power

Law of Practice, Speed-Accuracy-Trade-Off Functions), a decision was made to conduct all analyses separately for each of the three categories of performance measures. In other words, effect sizes from speed-based performance measures were not aggregated with effect sizes from accuracy-based performance measures – they were analyzed in separate meta-analyses.

Effect sizes indicating performance on the same task within the same performance measure category (e.g., number of sunk putts and average distance from the hole on a golf task) were averaged to obtain a single effect size. When the same subjects completed multiple tasks that represented procedural skills within the same performance measure category, effect sizes were aggregated if the skills that each task required were determined to be sufficiently similar (e.g., endotracheal intubation and chest tube insertion) but were treated independently if the skills for each task were determined to be sufficiently different (e.g., a balancing task vs. a tracking task). Effect sizes for the same participants were also treated independently when a study consisted of two retention intervals (e.g., participants were evaluated one month after training and again six months after training), and the effect size for the second retention interval was coded with a performance opportunity during the retention interval. While the lack of independence of some data points is not ideal, a decision was made to prioritize the number of longer retention intervals and the breadth of procedural skills in these rare circumstances⁷.

4.2 RQ 1: Comparison to Arthur et al. (1998) Findings for Procedural Skills

⁷ Results were computed with and without the non-independent data points and results were similar.

The first stage of analyses involved replicating the findings from the Arthur et al. (1998) meta-analysis given that it is the most well-known existing meta-analysis on skill retention⁸. This was done using only the sources gathered that were published in 1998 or earlier (the year of Arthur et al.'s publication and earlier). Although there are now more modern meta-analytic techniques available, the meta-analytic strategy used in that paper was replicated in order to demonstrate what the authors likely would have found had they focused only on procedural skills and used the same databases that were used in the current study. The authors leveraged formulas from Hunter and Schmidt (1990). Specifically, mean effect sizes for specific categories of retention intervals were calculated using Equation 3 below:

$$\bar{d} = \frac{\sum d_i n_i}{N_T} \quad (3)$$

Where d = mean effect size, d_i = effect size for each study, n_i = sample size for each study, and N_T = total sample size across all studies. A bias multiplier calculated based on Equation 4 below was then applied to correct for uneven sample sizes. The corrected d was calculated as the mean d divided by A .

$$A = 1 + \frac{0.75}{N - 3} \quad (4)$$

where N = the average sample size across studies

⁸ I chose to replicate the Arthur et al. (1998) rather than the Wang et al. (2013) meta-analysis for two reasons: (1) It is generally more well-known, and (2) it was published as a peer-reviewed journal article rather than as a chapter.

Sampling error variance was estimated based Equation 5:

$$\text{Sampling error variance} = \frac{N - 1}{N - 3} \times \frac{4}{N} \times \frac{\text{corrected } d^2}{8} \quad (5)$$

The population variance [Var (δ)] was then calculated as the variance of the observed effect sizes minus the estimated sampling error variance. Finally, the standard deviation of the population effect sizes (SD δ) was calculated based on Equation 6:

$$SD \delta = \frac{\text{Var} (\delta)}{A} \quad (6)$$

4.3 RQ 2: Rate of Decline When Procedural Skills Are Infrequently Used

The second stage included the primary analyses designed to answer the research questions of the current study with all sources published through October of 2022 (when the literature search was conducted). The analyses were conducted with the effect sizes computed using the formula for independent samples (i.e., the majority of the effect sizes coded). For each performance measure category, a random effects meta-regression equation with retention interval predicting effect size was computed. The beta coefficient corresponding to retention interval was interpreted as the change in effect size observed for each month that the retention interval increased. Sampling error variance was estimated using Equation 7, which is recommended for random effects meta-analyses (Bornstein et al, 2009):

$$vi = \frac{N_1 + N_2}{N_1 \times N_2} + \frac{d^2}{2 \times (N_1 + N_2)} \quad (7)$$

Between-study heterogeneity (τ^2) was estimated using Restricted Maximum Likelihood (REML). In addition to τ^2 , the heterogeneity statistics of Cochran's Q , I^2 and H^2 were reported for each meta-regression model.

4.4 RQ 3-8: Moderators of the Retention Interval – Effect Size Relationship

Cochran's Q statistic was used to test for residual heterogeneity in order to determine if there was sufficient heterogeneity to investigate moderators. A significant test of Cochran's Q statistic suggests significant heterogeneity in residuals across studies. Significant heterogeneity would indicate that there are other factors beyond retention interval that contribute to observed differences in effect sizes and would therefore justify moderator analyses. A lack of significance would indicate that there is not sufficient heterogeneity to justify moderator analyses. Moderators were then tested by entering variables representing each moderator into meta-regression equations predicting effect size separately and the significance of the beta coefficient for the interaction term (retention interval by moderator) was used to determine the significance of the moderator. The results of these moderator analyses indicated whether each moderator altered the relationship between retention interval and effect size magnitude, which is a step beyond moderator analyses in prior skill retention meta-analyses that only looked at differences in overall effect size magnitudes across levels of moderators and did not account for change over time at different levels of each moderator. Plots of observed effect sizes at each level of

moderators were also provided to aid with interpretation of the practical significance of each moderator.

4.5 RQ 9: Loss of Initial Skill Acquisition Gains

In the third stage of analyses, the results of the meta-regression were interpreted in the context of initial acquisition gains. For each performance measure category, effect sizes indicating performance differences between initial and final acquisition were subjected to a random-effects meta-analysis. The same parameters used in the meta-regression equation for the relationship between retention interval and effect size (e.g., sampling error variance estimation, REML) were used in the computation of the meta-analysis of acquisition data. The only difference was that rather than examining the influence of time on effect size, this meta-analysis examined acquisition effect sizes as a whole, split out by accuracy, speed, and mixed performance measures. This allowed for the comparison of initial acquisition effect sizes to retention effect sizes, which led to an estimation of the approximate proportion of skill acquisition gains that may be lost over intervals of non-use or infrequent use.

4.6 RQ 10: Comparison of Effect Sizes Computed with Different Formulas

Meta-regression equations with retention interval predicting effect size were computed using the effect sizes that were converted from paired samples *t*-values and reported as Cohen's *D* statistics. The same model parameters as were used in previous models were used (e.g., sampling error variance estimation, REML). The overall strength of effect sizes as well as the relationship between effect size magnitude and retention interval length were compared to the results of the models computed using the independent

samples formula effect sizes. There were not enough effect sizes calculated without the independent samples formula to conduct moderator analyses.

4.7 Publication Bias

Finally, the fourth stage of analyses involved an assessment of publication bias. In any meta-analysis, publication bias can confound results through the exclusion of outcomes or effect sizes that do not make it through to publication. Therefore, it is critical that meta-analysts not only take measures to prevent publication bias when conducting their literature search, but also statistically test for publication bias within the studies included in meta-analyses (Ferguson & Brannick, 2012). A two-pronged approach was taken to account for publication bias. First, measures were taken to prevent publication bias by including grey literature (i.e., theses, dissertations, conference papers, technical reports) in the literature review process. Publication bias was then directly evaluated within the studies included in the meta-analysis by testing a variable indicating whether each effect size came from published or grey literature as a moderator. The moderating effect of publication type was assessed to determine both the influence of publication time on effect size magnitude generally and the influence of publication type on changes in effect size magnitude over time. The moderating effect of publication type on effect size magnitude generally was tested using a random-effects meta-analysis for each performance measure category with subgroup analyses for each publication type. The moderating effect of publication type on the change in effect size magnitude over time was tested using a random-effects meta-regression with an interaction term for publication type by retention interval. The same model parameters (e.g., sampling error estimation, REML) previously specified for other

models were used. A significant moderating effect of literature type would suggest evidence of possible publication bias.

Second, the relationship between precision and effect size magnitude was evaluated in order to assess the possibility that publication bias may have prevented relevant findings from being included in the meta-analysis. Precision in the context of meta-analysis refers to the level of certainty or reliability associated with each effect size. Studies with higher levels of precision are characterized by large sample sizes and low standard errors. The relationship between precision and effect size magnitude is used to detect the presence of publication bias because if studies with lower levels of precision have stronger effect sizes, it would suggest the absence of small effect sizes with small samples from the published literature. This relationship is often assessed via funnel plots which plot effect size magnitude on the X-axis and standard error on the Y-axis (Stern & Egger, 2001). However, in the case of standardized mean differences, the effect size formula is used in the calculation of the standard error for each effect size, which can create an artificial association between precision and effect size magnitude and lead to misleading results (Lau et al., 2006). Because of this, based on the recommendations of Zetsloot et al. (2017), the square root of the sample size at retention was used as an index of precision for the current study.

An alternative to funnel plots, where effect size magnitude was plotted on the X-axis and the square root of the sample size was plotted on the Y-axis, was evaluated for the effect sizes in each performance measure category. The relationship between the two variables was also tested using Precision Effect Tests (PET) and Precision Effect Estimate with Standard Error (PEESE) tests (Stanley & Doucouliagos, 2014), which are regression-

based tests that evaluate the relationship between precision and effect size. Typically, PET uses the effect size standard error as an indicator of precision while PEESE uses sampling error variance as an indicator of precision, which provides a more robust estimate at high levels of heterogeneity. Because of the use of Cohen's D in these formulas, the square root of the sample size at retention and the sample size at retention were used as alternatives to standard error and sampling error variance, respectively. Asymmetry in the plots or significant PET-PEESE tests would suggest evidence for possible publication bias.

CHAPTER 5. RESULTS

In the Results chapter, I present all findings from meta-analytic models used to test the study's research questions. The 10 research questions that guided the study's design are used to structure the Results chapter. In the following pages, I address each research question individually by reporting the results of the analyses and the extent to which the results answer each research question. After addressing the results for each research question, I present the results of publication bias assessments and discuss the potential impact of publication bias on the results presented earlier in the chapter. As previously mentioned, the effect sizes computed using the independent samples formula were used to answer Research Questions 1-9 and the effect sizes reported by authors and converted from paired *t*-values are used to address Research Question 10.

5.1 RQ 1: Replication of Arthur et al. (1998) Findings for Procedural Skills

The first set of analyses involved a replication of the findings of Arthur et al. (1998) in order to determine what the authors may have found had they focused only on procedural skills with a significant motor component and used the academic databases that I did in my literature search. To obtain the results, only sources that were published in 1998 or earlier and satisfied the inclusion criteria for the current study were included and the meta-analytic strategy outlined by Arthur et al. (1998) was followed. Arthur et al.'s (1998) findings are reported in Table 2 and the replication results for accuracy performance measures, speed performance measures, and performance measures that involve a mixture of speed and accuracy are reported in Tables 3, 4, and 5, respectively.

The findings in this section are discussed with respect to the similarities and differences between the results reported by Arthur et al. (1998) and the results yielded from this replication. One similarity between the two meta-analyses is that the majority of performance data represented accuracy-based performance, rather than speed-based performance. Arthur et al. (1998) treated speed vs. accuracy as a moderator and their analyses included 145 accuracy-based effect sizes and 33 speed-based effect sizes (81% accuracy). The replication consisted of 199 accuracy-based effect sizes, 29 speed-based effect sizes, and 13 mixed effect sizes (82% accuracy-based effect sizes). Both analyses also suggest evidence of increasing skill decay over time, particularly at longer retention intervals. Results suggest that mean skill levels decline by at least one standard deviation unit in magnitude at retention intervals of six months or higher. For studies with retention intervals ranging from six months to one year, Arthur et al. reported an estimated Cohen's *D* of -1.03 [-1.80, -0.26] for all performance measures and the replication yielded an estimated Cohen's *D* of -1.12 [-2.40, 0.16] for accuracy-based performance measures. For studies with retention intervals over one year, Arthur et al. (1998) reported an estimated Cohen's *D* of -1.27 [-1.95, -0.58] across all performance measures and the replication yielded an estimated Cohen's *D* of -2.51 [-3.48, -1.54] for accuracy-based performance measures. It is worth noting that although both results showed substantial skill decay, the estimate beyond one year from the replication with only procedural skills was notably stronger. There were not sufficient data points to provide a meta-analytic estimate for speed-based or mixed performance measures in studies of retention intervals of six months or longer in the replication.

The most notable difference between the two sets of analyses pertains to the overall strength of effect sizes. Specifically, estimated effect sizes across all retention intervals appear notably greater decay in the Arthur et al. study ($\delta = -0.95 [-2.35, -0.44]$) compared to the current study ($\delta_{\text{Accuracy}} = -0.54 [-1.36, 0.28]$, $\delta_{\text{Speed}} = -0.23 [-1.12, 0.66]$, $\delta_{\text{Mixed}} = -0.36 [-1.49, 0.78]$). The differences are particularly stark at retention intervals of a week or less ($\delta_{\text{Arthur et al.}} = -1.01 [-2.53, 0.51]$, $\delta_{\text{Accuracy}} = -0.13 [-1.01, 0.75]$, $\delta_{\text{Speed}} = -0.12 [-1.31, 1.08]$). A second notable difference between the two analyses is the overall number of data points included in the two studies (178 in Arthur et al. vs. 241 in the current study). Potential explanations for these discrepancies are explored in the Discussion chapter.

It appears based on these results that the answer to Research Question 1 is that had Arthur et al. (1998) conducted their meta-analysis focused on procedural skills, rather than skills requiring both declarative and procedural knowledge, they likely would have obtained more modest effect sizes at shorter retention intervals. However, the general takeaway that longer retention intervals were associated with higher levels of decay would have remained consistent. Both Arthur et al.'s (1998) findings and the replication findings suggest the likelihood of substantial decay when skills are not used for a period of six months or more.

Table 2 – Arthur et al. (1998) Results

Retention Interval (Days)	Number of Data Points	Total Sample Size	Mean d	Corrected d	SD of population effect sizes	Variance Due to Sampling Error	Min d	Max d	Lower Bound	Upper Bound
Overall	178	8719	-0.97	-0.95	0.72	15.39%	-5.24	0.99	-2.35	0.44
Less than 1	3	45	0.02	0.02	0.00	100.00%	-0.12	0.17	0.02	0.02
1 to 7	89	3325	-1.03	-1.01	0.77	17.02%	-5.24	0.82	-2.53	0.51
8 to 14	13	680	-1.54	-1.52	0.63	20.20%	-2.87	-0.02	-2.75	-0.28
15 to 28	19	766	-0.95	-0.93	1.00	10.04%	-4.06	0.59	-2.90	1.03
29 to 90	33	2656	-0.65	-0.65	0.39	25.35%	-1.84	0.99	-1.42	0.12
91 to 180	7	302	-1.42	-1.39	0.82	14.83%	-3.33	-0.22	-3.00	0.22
181 to 365	11	670	-1.04	-1.03	0.39	32.89%	-2.67	-0.55	-1.80	-0.26
>365	3	275	-1.28	-1.27	0.35	30.19%	-2.48	-0.92	-1.95	-0.58

Note. ES = Effect size. Upper and lower bounds indicate 95% confidence intervals.

Table 3 – Replication of Arthur et al. Findings for Accuracy

Retention Interval (days)	Number of Data Points	Total Sample Size	Mean d	Corrected d	SD of population effect sizes	Variance Due to Sampling Error	Min d	Max d	Lower Bound	Upper Bound
Overall	196	5107	-0.56	-0.54	0.99	14.14%	-5.75	1.57	-1.36	0.27
7 or less	91	2004	-0.13	-0.13	0.56	37.03%	-2.26	1.57	-1.01	0.75
8 to 14	9	174	-0.78	-0.75	1.09	16.06%	-2.90	0.53	-1.72	0.23
15 to 28	14	1526	-0.66	-0.66	1.17	2.74%	-4.54	-0.25	-1.05	-0.27
29 to 90	67	1044	-0.75	-0.71	0.98	22.58%	-5.75	1.42	-1.81	0.39
91 to 180	2	61	-0.18	-0.18	0.00	100%	-0.25	-0.12	-0.91	0.56
181 to 365	6	79	-1.20	-1.12	0.61	50.07%	-2.18	-0.06	-2.40	0.16
>365	7	219	-2.58	-2.51	0.80	26.72%	-3.28	-0.76	-3.48	-1.54

Note. Note. ES = Effect size. Upper and lower bounds indicate 95% confidence intervals.

Table 4 – Replication of Arthur et al. Findings for Speed

Retention Interval	Number of Data Points	Total Sample Size	Mean d	Corrected d	SD of population effect sizes	Variance Due to Sampling Error	Min d	Max d	Lower Bound	Upper Bound
Overall	29	391	-0.24	-0.23	0.32	63.30%	-1.55	0.99	-1.12	0.66
7 or less	18	233	-0.13	-0.12	0.00	100.00%	-0.76	0.82	-1.31	1.08
8 to 14	3	36	-0.46	-0.42	0.65	45.91%	-1.55	0.20	-1.69	0.84
15 to 28	1	12	-0.70	-0.64	NA – 1 ES	NA – 1 ES	-0.70	-0.70	NA	NA
29 to 90	7	110	-0.38	-0.36	0.00	100.00%	-0.65	0.99	-1.43	0.72
91 to 180	0									
181 to 365	0									
>365	0									

Note. ES = Effect size. Upper and lower bounds indicate 95% confidence intervals.

Table 5 – Replication of Arthur et al. Findings for Mixed Performance

Retention Interval	Number of Data Points	Total Sample Size	Mean d	Corrected d	SD of population effect sizes	Variance Due to Sampling Error	Min d	Max d	Lower Bound	Upper Bound
Overall	13	186	-0.38	-0.36	0.78	32.36%	-2.19	0.80	-1.49	0.78
7 or less	2	48	0.69	0.67	0.00	100.00%	0.59	0.80	-0.19	1.53
8 to 14	2	12	-0.68	-0.55	0.00	100.00%	-0.50	-0.10	-2.65	1.56
15 to 28	0									
29 to 90	0									
91 to 180	7	101	-0.50	-0.46	0.62	43.47%	-2.19	-0.24	-1.60	0.67
181 to 365	1	17	-2.18	-2.07	NA – 1 ES	NA – 1 ES	-2.18	-2.18	NA	NA
>365	1	8	-1.07	-0.93	NA – 1 ES	NA – 1 ES	-1.07	-1.07	NA	NA

5.2 RQ 2: Rate of Decline When Procedural Skills Are Infrequently Used

The results presented for Research Question 1 relied solely on sources published prior to 1999 (because the Arthur et al. study was published the year prior), but the remaining results include sources published across all years. The eligibility screening and coding processes yielded a database of 1,281 effect sizes computed with the independent samples effect size formula representing 58,210 individuals at the time of acquisition and 30,364 individuals at the time of retention across 425 sources that was used to investigate the relationship between retention interval and procedural skill retention⁹. There were 493 retention intervals used across the 425 sources. A bar graph displaying the distribution of retention interval lengths within these 425 sources is presented in Figure 3. This distribution is characterized by a positive skew with a larger quantity of short retention intervals in comparison to long retention intervals. Over one-quarter (137, 28%) of the retention intervals were two days or shorter and nearly half (218, 44%) were shorter than one week. However, there were still a larger number of longer retention intervals than there were in previous meta-analyses of skill retention. Twenty-nine retention intervals were longer than six months and 14 retention intervals were longer than one year.

⁹ One study consisted of 13,176 participants at final acquisition, 726 participants at the first retention assessment, and 258 individuals at the second retention assessment. The discrepancy between sample sizes at acquisition and retention in this study accounts for a large portion of the discrepancy between the total sample sizes described in text. The lopsided samples do not substantially influence results because random-effects models account for asymmetric sample sizes when effect sizes are weighted by sampling error variance.

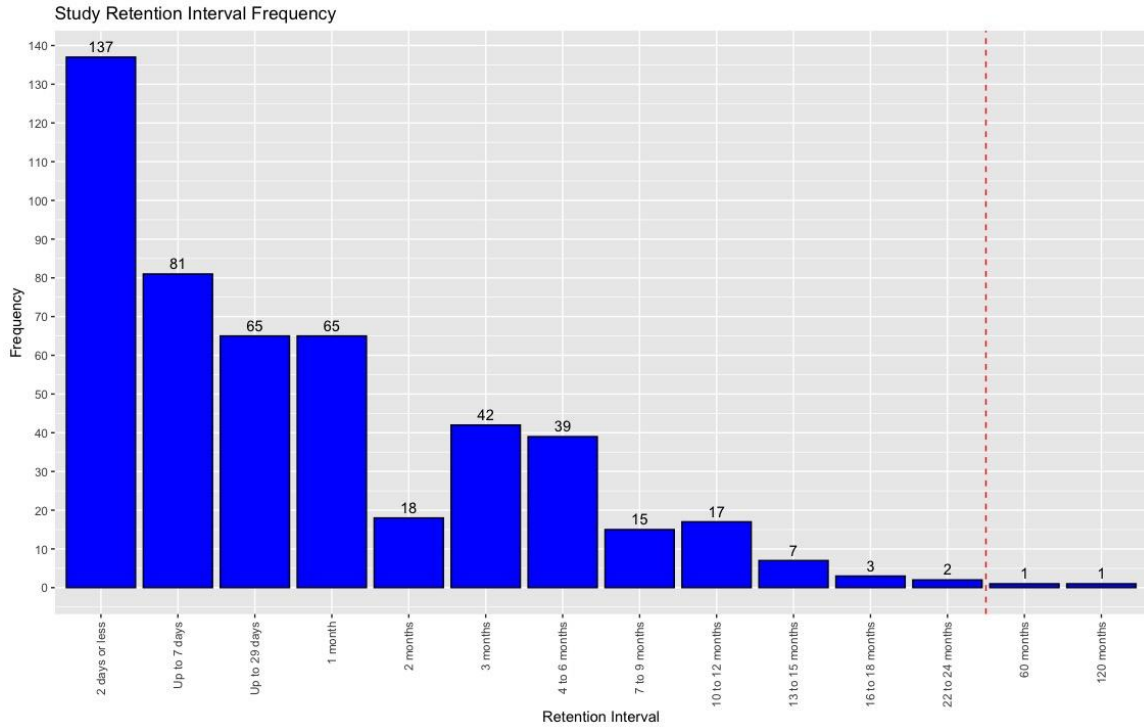


Figure 3 – Study Retention Interval Frequency

Note. Bar graph of study retention interval frequencies. The two outlier retention intervals of 60 and 120 months are separated in this graph by a red-vertical dotted line and were not included in meta-regression analyses.

Meta-regression models in which retention interval (on a metric of months) was used as a predictor of the effect size capturing the difference between performance at acquisition and performance at retention were computed to test the relationship between retention interval length and skill decline. Meta-regression allowed for the treatment of retention interval as a continuous variable. However, there were two large retention interval outliers – 60 months and 120 months (separated by the dotted red vertical line in Figure 3 and throughout visualizations in the Results chapter), which were both over twice as large as the next longest retention interval of 24 months. Because outliers this large would likely skew any estimation of the degree of skill decay that occurs per month, these studies were not included in meta-regression analyses but are reported in visualizations throughout the

results section to ensure all eligible data are represented. Separate meta-regression models were computed for accuracy-based performance measures, speed-based performance measures, and mixed performance measures. Scatterplots of observed effect sizes across retention intervals are also provided. Once the two studies with outlier retention intervals were removed, a total of 829 effect sizes across 311 sources were included in the accuracy meta-regression model, 267 effect sizes across 115 sources were included in speed meta-regression model, and 180 effect sizes across 90 sources were included in the mixed meta-regression model.

Figures 4, 5, and 6 include scatterplots of observed Cohen's *D* effect sizes across varying retention intervals. Larger data points in Figures 4, 5, and 6 (and throughout all figures reported in the results) indicate that they were given more weight in the meta-regression models (i.e., they had larger sample sizes and the sample sizes at acquisition and retention were relatively even). The three scatterplots suggest that skills generally decline between acquisition and retention, perhaps slightly more rapidly for accuracy-based performance measures than for speed-based or mixed performance measures. Additionally, the X-intercept of the mixed scatterplot appears lower than the X-intercept on the other two scatterplots which suggests that mixed performance measures displayed more rapid skill decay at earlier retention intervals.

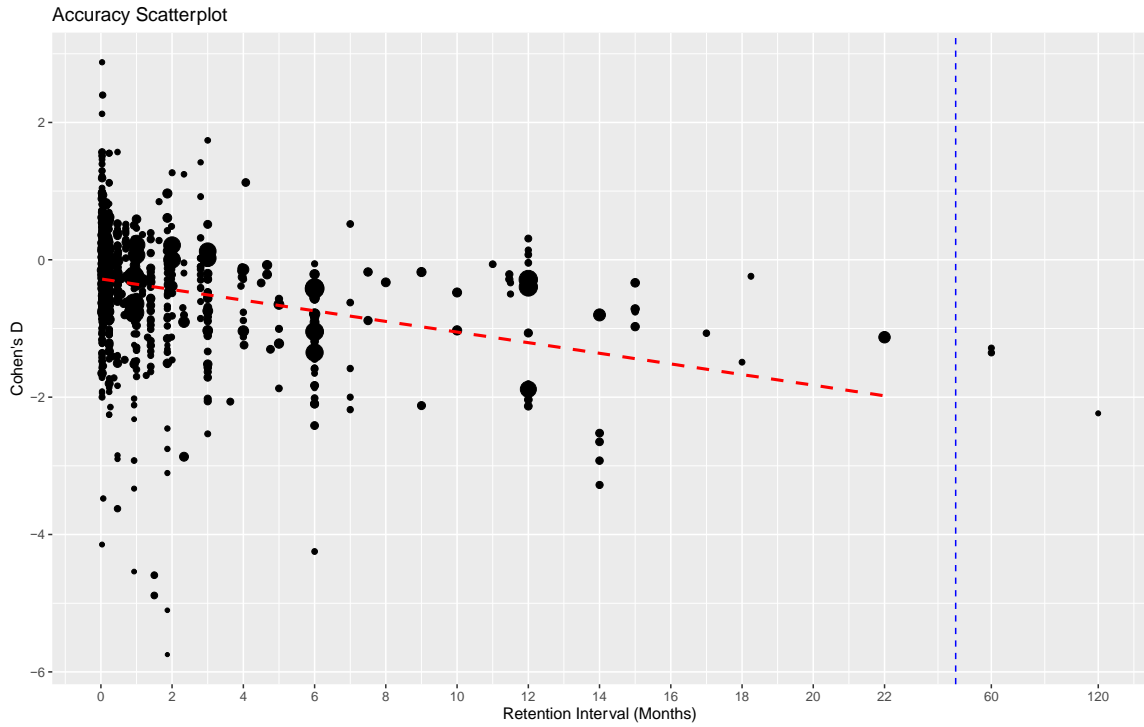


Figure 4 – Accuracy Scatterplot

Note. Scatterplot of effect size by retention interval for accuracy-based performance measures. Effect sizes that appear larger were given more weight in the meta-regression (i.e., they had lower sampling error variances). The two outlier retention intervals of 60 and 120 months are separated in this graph by a blue vertical-dotted line and were not included in meta-regression analyses, but are included here to ensure all eligible data are represented.

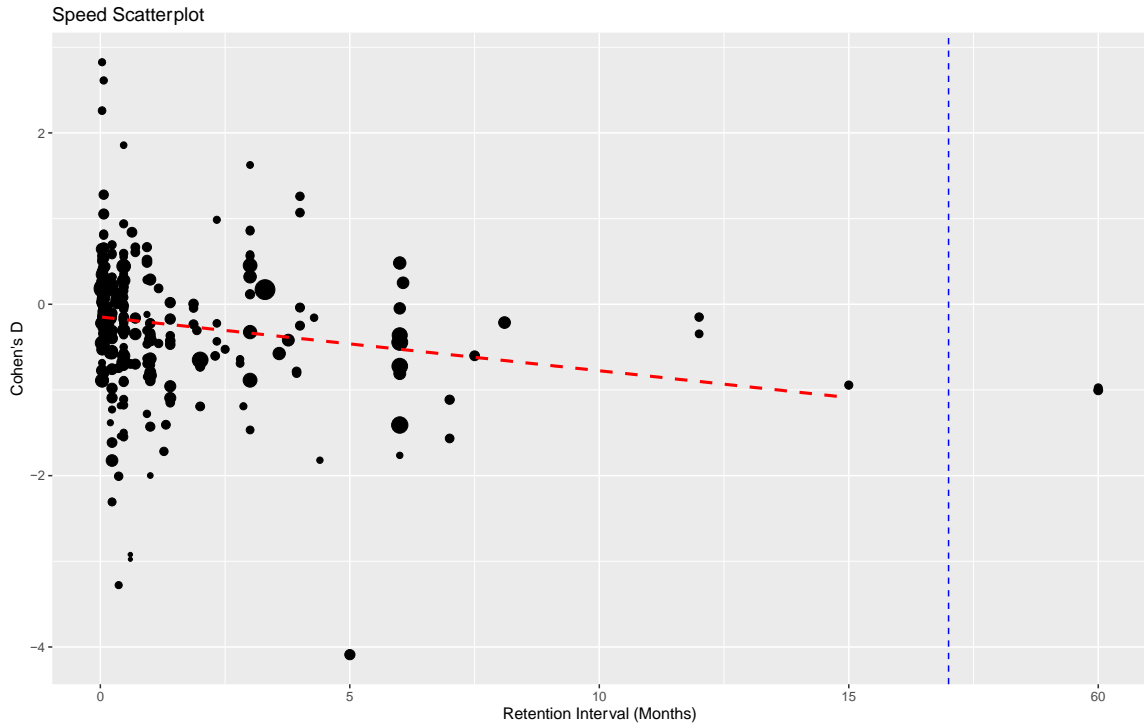


Figure 5 – Speed Scatterplot

Note. Scatterplot of effect size by retention interval for speed-based performance measures. Effect sizes that appear larger were given more weight in the meta-regression (i.e., they had lower sampling error variances). The two outlier retention intervals of 60 and 120 months are separated in this graph by a blue vertical-dotted line and were not included in meta-regression analyses, but are included here to ensure all eligible data are represented.

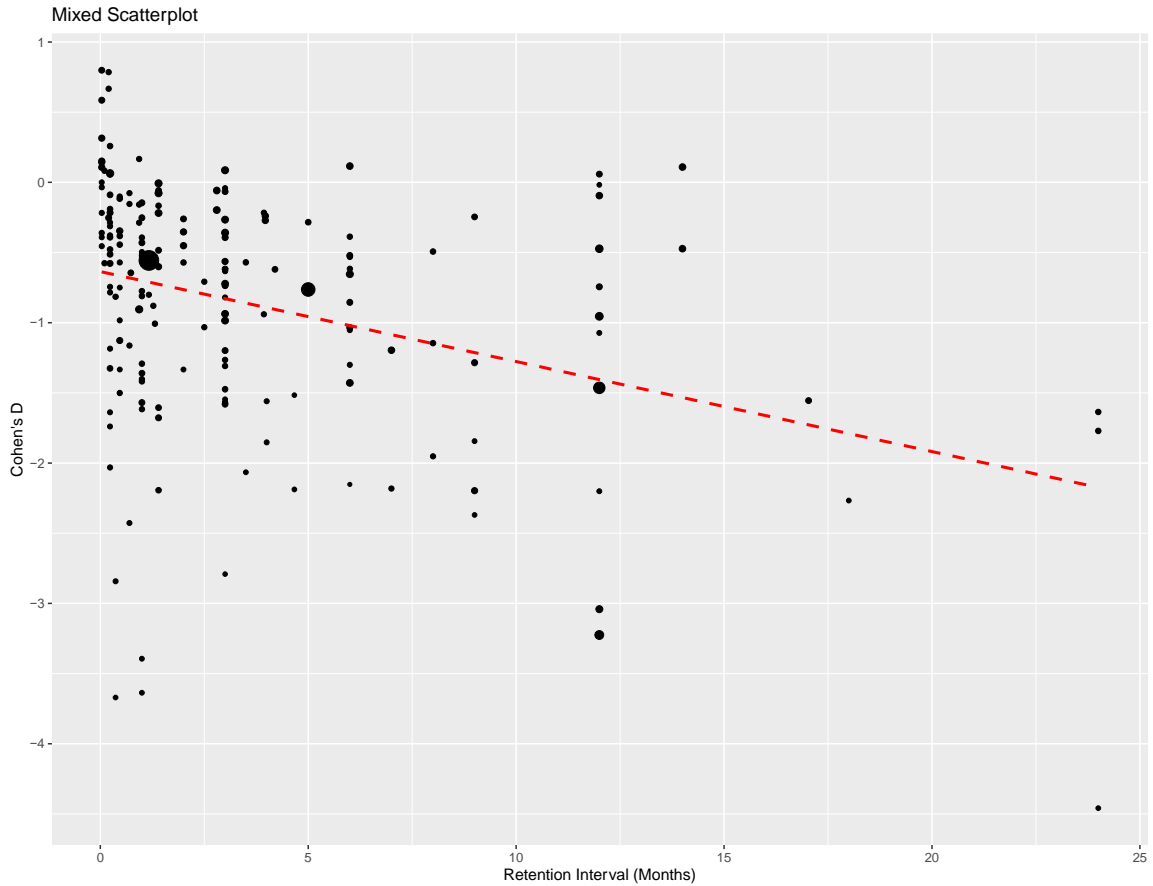


Figure 6 – Mixed Scatter Plot

Note. Scatterplot of effect size by retention interval for performance measures involving both speed and accuracy components. Effect sizes that appear larger were given more weight in the meta-regression (i.e., they had lower sampling error variances). The two outlier retention intervals of 60 and 120 months are separated in this graph by a blue vertical-dotted line and were not included in meta-regression analyses, but are included here to ensure all eligible data are represented.

Results from the meta-regression models confirmed the existence of a linear decline in skill proficiency levels across all three performance measure types¹⁰ (see Table 6 for the accuracy model, Table 7 for the speed model, and Table 8 for the mixed model). The beta

¹⁰ Non-linear functions were tested as well. They did not provide improvement in fit to the data for speed and mixed performance measures. While non-linear trends did provide an improvement in fit for accuracy performance measures, the non-linear trends revealed functions that did not decrease monotonically, which is not theoretically or practically justifiable in the context of skill retention. Therefore, a linear function was reported.

coefficient in the meta-regression model for accuracy performance measures suggests that the Cohen's D statistics indicating differences in performance between final acquisition and retention generally increases in magnitude by 0.08 per month of infrequent use ($\beta = -0.08$ [-0.10, -0.06]). The beta coefficients in the meta-regression models for speed and mixed performance suggest that Cohen's D statistics indicating differences in performance between final acquisition and retention increase in magnitude by approximately 0.06 per month ($\beta_{\text{Speed}} = -0.06$ [-0.11, -0.02], $\beta_{\text{Mixed}} = -0.06$ [-0.09, -0.04]). The beta-coefficients in all three models are significantly different from zero ($p < 0.01$), but the slightly stronger beta-coefficient for accuracy is not statistically different from the other two beta-coefficients, as evidenced by overlapping confidence intervals.

While a decline of 0.06 to 0.08 per month may not seem substantial on the surface, the magnitude of decline is more apparent when considered over longer periods of time. Table 9 presents the Cohen's D statistics that are predicted by the meta-regression models (i.e., based on the intercepts and the beta-coefficients) for each type of performance measure across increasing three-month intervals. According to these results, mean skill levels decline by a full standard deviation unit after 9 to 12 months of infrequent use for accuracy-based skills, 12 to 15 months of infrequent use for speed-based skills, and 6 to 9 months of infrequent use for skills that involve a mix of speed and accuracy. Mean accuracy-based and mixed skill levels decline by up to two standard deviation units after two years of infrequent use. The practical significance of these findings can be illustrated by converting a loss of one standard deviation unit to percentile loss based on a normal distribution. For example, consider individuals at the 90th and 50th percentiles in skill proficiency following skill acquisition. For a person initially at the 90th percentile,

infrequent use results in a decline to approximately the 61st percentile of final acquisition performance after 9 to 12 months for accuracy-based skills, 12 to 15 months for speed-based skills, and 6 to 9 months for mixed skills. A similar decline of a standard deviation unit for an individual initially at the 50th percentile would result in a decline to approximately the 16th percentile of final acquisition performance for each of the respective skill categories over the same durations. These results provide an answer to Research Question 2 through a quantitative estimate of the rate of skill decline that can be expected across intervals of infrequent use.

While the meta-regression results indicate that retention interval significantly predicts effect size, there is heterogeneity in effect sizes even after the variance attributed to retention interval is accounted for. A common rule of thumb for I^2 , a heterogeneity statistic interpreted as the proportion of total variation across studies that is due to true heterogeneity, is that I^2 values below 50% indicate low heterogeneity, between 50% and 75% indicate moderate heterogeneity, and 75% or higher indicate substantial heterogeneity (Higgins et al., 2003). Based on these thresholds, there is substantial heterogeneity in two of the three performance measures (83.21% for accuracy and 85.77% for mixed) even after accounting for retention interval and the third performance measure (speed) is approaching the threshold that separates moderate from substantial heterogeneity (72.50%). Cochran's Q test also indicated significant residual heterogeneity in all three meta-regression models ($Q_{\text{Accuracy}} = 4242.57, p < 0.01, Q_{\text{Speed}} = 907.02, p < 0.01, Q_{\text{Mixed}} = 1046.15, p < 0.01$), meaning that there is still unaccounted for effect size variation in each meta-regression, even after eliminating variance accounted for by retention interval length. These heterogeneity statistics indicate that an investigation into moderators is justified and may

reveal insight into factors that interact with length of retention interval to determine the strength of effect sizes between final acquisition and retention.

Table 6 – Accuracy Meta-Regression Model

Accuracy (k = 829)				
Model	β	<i>SE</i>	<i>t</i>	95% CI
Intercept	-0.25	0.03	-8.06**	[-0.31, -0.19]
Retention Interval	-0.08	0.01	-8.74**	[-0.10, -0.06]
Tests				
Residual Heterogeneity	Moderators			
$Q(827) = 4242.57^{**}$	$F(1, 827) = 76.33^{**}$			
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.42	0.65	83.21%	5.96	13.80%

Note. * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes. The test of residual heterogeneity is Cochran’s Q test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size.

Table 7 – Speed Meta-Regression Model

Speed (k = 267)				
Model	β	<i>SE</i>	<i>t</i>	95% CI
Intercept	-0.14	0.05	-2.53*	[-0.24, -0.03]
Retention Interval	-0.06	0.02	-2.89**	[-0.11, -0.02]
Tests				
Residual Heterogeneity	Moderators			
$Q(265) = 907.02^{**}$	$F(1, 265) = 8.35^{**}$			
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.35	0.59	72.50%	3.64	4.85%

Note. * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes. The test of residual heterogeneity is Cochran’s Q test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size.

Table 8 – Mixed Meta-Regression Model

Mixed (k = 180)				
Model	β	<i>SE</i>	<i>t</i>	95% CI
Intercept	-0.58	0.07	-7.87**	[-0.73, -0.44]
Retention Interval	-0.06	0.01	-4.88**	[-0.09, -0.04]
Tests				
Residual Heterogeneity	Moderators			
Q (178) = 1046.15**	F (1, 178) = 23.35**			
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.44	0.66	85.77%	7.03	15.43%

Note. * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes. The test of residual heterogeneity is Cochran's *Q* test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size.

Table 9 – Predicted Effect Sizes Based on Meta-Regressions

Retention Interval	Accuracy <i>D</i>	Speed <i>D</i>	Mixed <i>D</i>
3 months	-0.48	-0.33	-0.77
6 months	-0.72	-0.52	-0.96
9 months	-0.95	-0.71	-1.14
12 months	-1.18	-0.90	-1.33
15 months	-1.42	-1.09	-1.51
18 months	-1.65	-1.28	-1.70
21 months	-1.88	-1.47	-1.89
24 months	-2.12	-1.66	-2.07

Note. Effect sizes predicted by regression models (i.e., based in intercepts and beta coefficients) for increasing retention interval increments of three months up to two years.

5.3 Moderator Analyses

The task-related variables and study design variables included in the coding protocol were tested as moderators of the relationship between retention interval and effect size between final acquisition and retention. Task-related moderators included laboratory/artificial vs. real-world tasks, broad task type (i.e., medical/dental, military and

transportation, sports, laboratory, miscellaneous), open-looped vs. closed-loop, complexity (i.e., component, coordinative, dynamic), and ability demands (i.e., dexterity, gross body coordination, precision/steadiness). Study design moderators included opportunities for infrequent performance during the retention interval and task instructions.

In order to test for variables that moderate the relationship between retention interval and effect size between final acquisition and retention, each of the proposed moderators was entered separately into a meta-regression equation similar to those previously reported (i.e., with retention interval predicting effect size between final acquisition and retention). The extent to which each variable altered the relationship between retention interval and effect size was determined by beta coefficients for interaction terms between the moderator and the retention interval, which are interpreted as the difference in slope of the regression line between retention interval and effect size between the reference group and other levels of the moderator (for categorical moderators) or as the change in the regression line per unit increase of the moderator (for continuous moderators). The statistical significance of moderator effects was determined by the *p*-value associated with each interaction term. Beta coefficients for interaction terms and 95% confidence intervals associated with each moderator are reported in Table 10. For the purpose of brevity, only the beta coefficients for the interaction terms, as opposed to the full model statistics, are presented in this section. However, full meta-regression models for each moderator tested can be found in Appendices C-M¹¹. Scatterplots that are color-

¹¹ Sample age, sample type, number of acquisition trials, and acquisition effect size were also tested as moderators, but the results were not illuminating and are not reported for the sake of brevity.

coded based on the level of each moderator are provided throughout the section to aid with interpretation.

Results are presented for each moderator, including its effect on accuracy, speed, and mixed performance measures. In many instances, moderating effects that were statistically significant for accuracy were not statistically significant for speed, but scatterplots revealed similar trends across the two performance measure types. The lack of statistical significance for speed-based performance measures in these instances may have been an issue of power for three reasons: First, there were over four times more effect sizes associated with accuracy-based performance measures than effect sizes associated with speed-based performance measures (829 vs. 267). Second, speed was the only performance measure type with an I^2 below the typical threshold for high heterogeneity (75%) and there may not have been enough heterogeneity in effect sizes to detect some meaningful moderating effects. Third, power to detect moderating effects may have been particularly low for speed-based performance measures because it was characterized by the smallest number of long retention intervals. Only 69 of the effect sizes in of the effect sizes for speed-based performance measures (26%) had retention intervals longer than one month. Accuracy-based performance measures had a larger overall number of data points, and mixed performance had a larger proportion of retention intervals over one month (56%), both of which resulted in increased power to detect moderating relationships compared to speed-based performance measures. There were also a few cases where moderators were not significant for mixed performance measures, but scatterplots revealed potentially meaningful trends. Moderator analyses are reported both with respect to statistical and

practical significance, but results lacking statistical significance should accordingly be interpreted with caution and considered suggestive rather than definitive.

Table 10 – Moderator Analyses

	Accuracy	Speed	Mixed
<i>Task-Related Moderators</i>			
Lab ^R vs. Real World Task	0.11** [0.06, 0.15]	0.07 [-0.04, 0.19]	
Task Type (Laboratory/Artificial ^R)			
Medical/Dental	0.12** [0.07, 0.17]	0.08 [-0.03, 0.20]	
Military & Transportation	0.16** [0.07, 0.24]	-0.03 [-0.24, 0.17]	
Sports	0.18** [0.09, 0.27]	-2.05 [-7.97, 3.87]	
Miscellaneous	0.13** [0.06, 0.21]	0.42** [0.14, 0.70]	
Open vs. Closed-Looped ^R	0.01 [-0.03, 0.04]	0.06 [-0.03, 0.14]	0.11** [0.05, 0.16]
Complexity			
Component Complexity	0.03** [0.01, 0.05]	0.02 [-0.03, 0.07]	0.06** [0.03, 0.09]
Coordinative Complexity	0.02 [-0.01, 0.04]	0.01 [-0.05, 0.06]	0.07** [0.03, 0.11]
Dynamic Complexity	0.02 [0.00, 0.05]	0.04 [-0.01, 0.09]	0.06** [0.03, 0.09]
Abilities (No ^R)			
Dexterity	0.10 [-0.08, 0.27]	-0.08 [-0.27, 0.10]	-0.25* [-0.49, -0.02]
Gross Body Coordination	0.09* [0.01, 0.18]	0.02 [-0.15, 0.20]	0.02 [-0.16, 0.19]
Precision/Steadiness	0.01 [-0.03, 0.04]	0.01 [-0.08, 0.10]	0.10** [0.03, 0.16]
<i>Study Design Moderators</i>			
Performance Opportunities (No ^R)	0.06** [0.02, 0.11]	0.02 [-0.08, 0.13]	0.06 [0.00, 0.13]

Task Instructions		
Accuracy ^R		
Speed or Mixed	-0.08* [-0.16, -0.01]	
Not Clear	0.04* [0.00, 0.08]	
Speed ^R		
Accuracy or Mixed		-0.11 [-0.26, 0.05]
Not Clear		-0.04 [-0.19, 0.12]
Mixed ^R		
Speed or Accuracy		-0.22 [-0.50, 0.05]
Not Clear		-0.01 [-0.10, 0.08]

Note. * < 0.05, ** < 0.01. ^R = Reference group. Cells indicate beta coefficients for the interaction between retention interval and each moderator. Beta coefficients for each combination of moderator and performance measure were computed in separate meta-regression mode

5.3.1 *Task-Related Moderators*

5.3.1.1 RQ 3: Task Type as a Moderator

Task type was only tested as a moderator for speed and accuracy-based performance measures because of a lack of task-type diversity in the mixed performance measure category. Over 80% (150 out of 180) of the effect sizes in the mixed performance measure category involved medical/dental tasks. The moderating effect of task type was tested in two steps. First, a binary variable indicating whether the task was a laboratory/artificial task or a real-world task was tested. There were 554 (67%) effect sizes representing real-world tasks and 275 (33%) effect sizes representing laboratory/artificial tasks in the accuracy model and there were 185 (69%) effect sizes representing real-world tasks and 82 (31%) effect sizes representing laboratory/artificial tasks in the speed model. Results indicated that this variable was a statistically significant moderator for accuracy-based performance measures but not for speed-based performance measures. The same trend was observed for speed, although the results were not statistically significant. The slope of the regression line between retention interval and effect size was shallower (i.e., the relationship was weaker) for real-world tasks than for laboratory/artificial tasks, which indicates that performance on laboratory tasks declined faster over time than performance on real-world tasks did ($\beta_{\text{Accuracy}} = 0.11, \beta_{\text{Speed}} = 0.07$, see Figure 7).

Second, more granular task-type categories (i.e., laboratory/artificial, medical/dental, military and transportation, sports, and miscellaneous) were tested as a moderator. While it was originally planned that military and transportation tasks would be included as separate categories, there were relatively few studies in each category and the categories were collapsed because the skills required by tasks in each category were deemed to be sufficiently similar. Among effect

sizes in the accuracy model, 253 (31%) were from medical/dental tasks, 208 (25%) were from sports tasks, 66 (8%) were from military and transportation tasks, and 27 were from miscellaneous tasks (3%). For the speed model, 108 (40%) were from medical/dental tasks, 9 (3%) were from sports tasks, 22 (8%) were from military and transportation tasks, and 46 (17%) were from miscellaneous tasks. Laboratory/artificial tasks (275 or 33% of accuracy-based effect sizes and 82 or 31% of speed-based effect sizes) were chosen as the reference group in order to determine whether performance on all real-world tasks is better retained than performance on laboratory/artificial tasks or whether there are some real-world task categories that are better retained in comparison to laboratory/artificial tasks.

Results indicated that task type was once again a statistically significant moderator of the relationship between retention interval and effect size for accuracy-based performance measures but not for speed-based performance measures. For accuracy-based performance measures, slopes indicating the relationship between retention interval and effect size were stronger for laboratory/artificial tasks than they were for all real-world task types ($\beta_{\text{Medical/Dental}} = 0.12$, $\beta_{\text{Military/Transportation}} = 0.16$, $\beta_{\text{Sports}} = 0.18$, $\beta_{\text{Miscellaneous}} = 0.13$), which suggests that performance on laboratory/artificial tasks declines faster over time than it does for all categories of real-world tasks. Analyses for speed-based performances showed the same trend for medical/dental and miscellaneous tasks ($\beta_{\text{Medical/Dental}} = 0.08$, $\beta_{\text{Miscellaneous}} = 0.42$), but revealed stronger slopes for military and transportation ($\beta = -0.03$) and sports tasks ($\beta = -2.05$). However, the speed-based results for military and transportation tasks should be interpreted with caution because they are based on a smaller number of observations (9 sports effect sizes and 22 military/transportation effect sizes, in comparison to 108 medical/dental effect sizes and 46 miscellaneous effect sizes). See Figure 8 for a visualization of the moderating effect of all task types.

These results suggest that the answer to Research Question 3 is that task type does in fact appear to moderate the relationship between retention interval length and skill decay, at least for accuracy-based performance measures and with non-significant trends for speed in the same direction. The moderating effects are mostly attributed to differences between laboratory vs. artificial tasks. Performance on real-world tasks was better retained over time than performance on laboratory/artificial tasks for accuracy-based performance measures. There is less evidence, however, that skill decay patterns differ between different types of real-world tasks (i.e., medical/dental, military/transportation, sports).

5.3.1.2 RQ 4: Open-Looped vs. Closed-Looped Task as a Moderator

There were 359 (43%) effect sizes representing open-looped tasks compared to 470 (57%) effect sizes representing closed-looped tasks in the accuracy model, 138 (52%) representing open-looped tasks compared to 129 (48%) representing closed-looped tasks in the speed model, and 114 (63%) representing open-looped tasks compared to 66 (37%) representing closed-looped tasks in the mixed model. Closed-looped tasks was chosen as the reference group. Results indicated that open vs. closed loop tasks was a statistically significant moderator of the relationship between retention interval and effect size for mixed performance measures, but not for accuracy or speed-based performance measures. The direction of the moderating effect for mixed performance measures indicated that open-looped tasks had a shallower regression line slope for the relationship between retention interval and effect size than closed-looped tasks ($\beta = 0.11$, $p < 0.01$), which suggests that performance on open-looped tasks was better retained over time than performance on closed-looped tasks (see Figure 9). The same trend was observed for speed-based performance measures, but parallel regression lines in the accuracy plot in Figure 9 indicate that retention trends

are similar for open-looped vs. closed-loop accuracy tasks for accuracy-based performance measures.

These results suggest somewhat mixed results for the answer to Research Question 4. On one hand, open vs. closed-looped tasks do appear to moderate the relationship between retention interval and effect size for speed-based and mixed performance measures such that performance on closed-looped tasks decays faster than performance on open-looped tasks (although the trend for speed was not statistically significant). However, open vs. closed-looped tasks do not appear to moderate the relationship for accuracy-based performance measures. Potential explanations for these mixed results are further explored in the Discussion chapter.

5.3.1.3 RQ 5: Task Complexity as a Moderator

Generally speaking, descriptive statistics from task complexity scores associated with effect sizes in each performance measure category indicate that, on average, effect sizes in the accuracy (Component: $M = 1.64$, $SD = 0.71$, Coordinative: $M = 1.99$, $SD = 0.73$, Dynamic: $M = 1.51$, $SD = 0.71$) and speed-based (Component: $M = 1.81$, $SD = 0.90$, Coordinative: $M = 2.19$, $SD = 0.80$, Dynamic: $M = 1.79$, $SD = 0.85$) performance measure categories were associated with low to moderate complexity tasks (with the exception of coordinative complexity for speed) and effect sizes associated with mixed performance measure categories were associated with moderate to high complexity tasks (Component: $M = 2.21$, $SD = 0.88$, Coordinative: $M = 2.50$, $SD = 0.76$, Dynamic: $M = 2.28$, $SD = 0.82$). Component complexity was a statistically significant moderator of the relationship between retention interval and effect size for accuracy-based and mixed performance measures ($\beta_{\text{Accuracy}} = 0.03$, $p < 0.01$; $\beta_{\text{Mixed}} = 0.06$, $p < 0.01$). Coordinative and dynamic complexity were statistically significant moderators of the relationship between retention

interval and effect size for mixed performance only ($\beta_{\text{Coordinative}} = 0.07, p < 0.01; \beta_{\text{Dynamic}} = 0.06, p < 0.01$). Complexity dimensions were not statistically significant moderators of the relationship between retention interval and effect size for speed-based performance measures. While some of the effects were small or non-significant, there was a general pattern of better retention over time for high-complexity tasks in comparison to low-complexity tasks across all three dimensions of task complexity, as indicated by positive beta-coefficients across interaction terms in the meta-regression equations for all three performance measure categories. In other words, performance on high-complexity tasks appears to be better retained over time than performance on low-complexity tasks.

The results for the complexity moderator analyses suggest that the answer to Research Question 5 is that task complexity may in fact moderate the relationship between retention interval and skill retention, particularly for mixed performance measures which had statistically significant results for all three complexity dimensions. Further, component complexity may be a dimension of complexity that is a particularly meaningful moderator because it was significant for two of the three performance measures. The overall pattern of results (positive interaction terms) suggests that task complexity as a whole may be an important moderator for all performance measures, but the lack of statistical significance for five of the six speed and accuracy interaction terms prevents a definitive conclusion.

5.3.1.4 RQ 6: Task Ability Demands as a Moderator

For all ability demand moderator analyses, “No” was chosen as the reference group, meaning that beta coefficients for interaction terms indicate the difference in slope of the regression line for the relationship between retention interval and performance for tasks that

require each ability in comparison to tasks that do not require each ability. There were high and low base rates associated with whether the tasks included in each performance measure category required dexterity. A majority of all effect sizes represented tasks that involved dexterity (711 or 86% of accuracy-based effect sizes, 234 or 88% of speed-based effect sizes, and 176 or 98% of mixed effect sizes). Dexterity was a statistically significant moderator for mixed performance measures, such that performance on tasks involving dexterity declines faster over time than it does on tasks not involving dexterity ($\beta = -0.25, p < 0.01$). The direction of beta coefficients indicates that speed-based and mixed performance declined faster over time when tasks involved dexterity compared to when they did not involve dexterity, but that accuracy-based performance declined faster over time when tasks did not involve dexterity compared to when they did involve dexterity ($\beta_{\text{Accuracy}} = 0.10, \beta_{\text{Speed}} = -0.08, \beta_{\text{Mixed}} = -0.25$). However, these results should be interpreted with caution because of the extremely low base rates.

There were also low base rates associated with whether tasks included in the speed and mixed performance measure categories involved gross body coordination. A very small number of effect sizes in the speed and mixed categories represented tasks that required gross body coordination (36 or 13% of speed-based effect sizes, 5 or 3% of mixed effect sizes). However, there was a larger number of effect sizes from accuracy-based performance measures that represented tasks that required gross body coordination (235 or 40%) and gross body coordination was a statistically significant moderator of the relationship between retention interval and effect size for accuracy-based performance measures. The regression line indicating the relationship between retention interval and effect size was steeper for tasks that did not involve gross body coordination than for tasks that did involve gross body coordination ($\beta = 0.09, p < 0.05$), which indicates that performance on tasks that involved gross body coordination was better retained over

time than performance on tasks that do not involve gross body coordination (see Figure 10). However, only five effect sizes involved tasks that required gross body coordination and had retention intervals of two months or longer. Performance on tasks that involve gross body coordination was also slightly better retained for speed and mixed performance measures, but the magnitude of differences was small and lacked statistical significance and there were extreme base rates with these categories ($\beta_{\text{speed}} = 0.02, ns$; $\beta_{\text{Mixed}} = 0.02, ns$).

There were 562 (67%) effect sizes for accuracy-based performance measures that represented tasks that required precision/steadiness, 156 (58%) effect sizes for speed-based performance measures that represented tasks that required precision/steadiness, and 137 (76%) effect sizes for mixed performance measures that represented tasks that required precision/steadiness. Precision/steadiness was a statistically significant and meaningful moderator only for mixed performance measures. The regression line indicating the relationship between retention interval and effect size was steeper for tasks that did not involve precision/steadiness than for tasks that did involve precision/steadiness ($\beta = 0.10, p < 0.05$), which suggests that performance on tasks that require precision/steadiness was better retained over time than performance on tasks that do not require precision/steadiness was (see Figure 11). Precision/steadiness was not a statistically significant or meaningful moderator for accuracy or speed-based performance measures ($\beta_{\text{Accuracy}} = 0.01, ns$; $\beta_{\text{Speed}} = 0.01, ns$).

The results of these analyses unfortunately do not provide evidence that can be used to confidently conclude an answer to Research Question 6. The results suggest that it is possible that tasks involving gross body coordination, at least for accuracy-based performance measures, and tasks involving precision/steadiness, at least for mixed performance measures, may be better retained than tasks that do not require these abilities. However, the extremely low base rates

associated with most of these analyses cast doubt on the findings. The results may be suggestive, but they are not sufficient to conclude that task ability demands meaningfully moderate the relationship between retention interval and skill decline. Therefore, Research Question 6 remains somewhat of an open question.

5.3.2 *Study Design Moderators*

5.3.2.1 RQ 7: Performance Opportunities as a Moderator

Infrequent performance opportunities during the retention interval were reported for 83 (10%) of the effect sizes in the accuracy model, 29 (11%) of the effect sizes in the speed model, and 32 (18%) of the effect sizes in the mixed performance measure model. “No” was chosen as the reference group for performance opportunities, meaning that the beta coefficient associated with the interaction term is interpreted as the change in the slope of the regression line capturing the relationship between retention interval and effect size when there are infrequent opportunities during the retention interval in comparison to when there are not opportunities during the retention interval.

Performance opportunities during the retention interval were a statistically significant moderator of the relationship between retention interval and effect size for accuracy-based performance measures ($\beta_{\text{Accuracy}} = 0.06, p < 0.01$). For both accuracy-based and mixed performance measures, a general pattern (though not statistically significant for mixed performance measures) was observed in which there was less decay over time when infrequent performance opportunities occurred during the retention interval ($\beta_{\text{Mixed}} = 0.06, ns$). However, this pattern was not observed for speed-based performance measures ($\beta_{\text{Speed}} = 0.02, ns$). See Figure 12

for a visualization of the moderating effect of infrequent performance opportunities during the retention interval on the relationship between retention interval and effect size.

Based on this evidence, it appears that there is somewhat mixed evidence for Research Question 7. Results suggest that performance opportunities did moderate the relationship between retention interval and skill retention for accuracy-based performance measures, may have for mixed performance measures as well (due to the pattern of results but lack of significance), and did not meaningfully moderate for speed-based performance measures. Explanations for null findings with respect to speed are explored in the discussion section, but given the results for the other two performance measure categories, it is likely that performance opportunities are a meaningful moderator of skill decline rate.

5.3.2.2 RQ 8: Task Instructions as a Moderator

For task instruction moderator analyses, the reference group was chosen based on the extent to which instructions aligned with the performance measure. The reference group for each meta-regression equation was the task instruction category that aligned with the performance measure (i.e., accuracy instructions for accuracy-based performance measures, speed instructions for speed-based performance measures, both accuracy and speed instructions for mixed performance measures). Therefore, beta coefficients associated with each interaction term are interpreted as the change in slope of the regression line capturing the relationship between retention interval and effect size when task instructions were not aligned with the task performance measure or they were not clearly reported, compared to when they were aligned with the task performance measure.

For accuracy-based performance measures, 514 (62%) effect sizes involved only accuracy-based instructions, 103 (12%) involved speed-based or mixed instructions, and 212 (26%) effect sizes did not have clearly reported task instructions. For speed-based performance measures, 35 (13%) of effect sizes involved only speed-based instructions, 112 (42%) involved accuracy or mixed instructions, and 120 (45%) effect sizes did not have clearly reported task instructions. For mixed performance measures, 49 (27%) effect sizes involved both speed and accuracy-based instructions, 7 (4%) involved just speed or accuracy-based instructions, and 124 (69%) effect sizes did not have clearly reported task instructions.

Task instructions was a statistically significant moderator of the relationship between retention interval and effect size for accuracy-based performance measures. The beta coefficient associated with speed or mixed instructions was negative ($\beta_{\text{Speed or mixed instructions}} = -0.08, p < 0.05$), which indicates that performance was better retained over time when task instructions were aligned with how performance was scored (i.e., accuracy instructions for accuracy-based performance measures). Although the other effects were not statistically significant, the same general pattern was observed for speed and mixed performance measures. That is, the beta coefficients associated with the interaction term for mismatched instructions were negative ($\beta_{\text{Speed performance, accuracy or mixed instructions}} = -0.11, ns$; $\beta_{\text{Mixed performance, speed or accuracy instructions}} = -0.22, ns$), which indicates that performance is generally better retained when task instructions are aligned with how performance is measured (see Figure 13). The overall pattern of results and the significant findings for accuracy-based performance measures suggest that the answer to Research Question 8 is that task instructions did in fact moderate the relationship between retention interval and skill retention, although some of the results were not statistically significant.

In summary, the results of moderator analyses indicate that laboratory vs. real-world tasks, task complexity, performance opportunities, and task instructions may be meaningful moderators of skill decline. There may be other moderators (e.g., open vs. closed-looped tasks, gross body coordination demands, precision/steadiness demands), but the evidence is not as strong as it is for the former list of moderators. However, in many cases, inconsistent results across performance measures, a lack of statistical significance, or extreme base rates qualify the strength of the evidence and the results should accordingly be considered suggestive rather than definitive. A narrative summary of all moderator analyses is provided in Table 11.

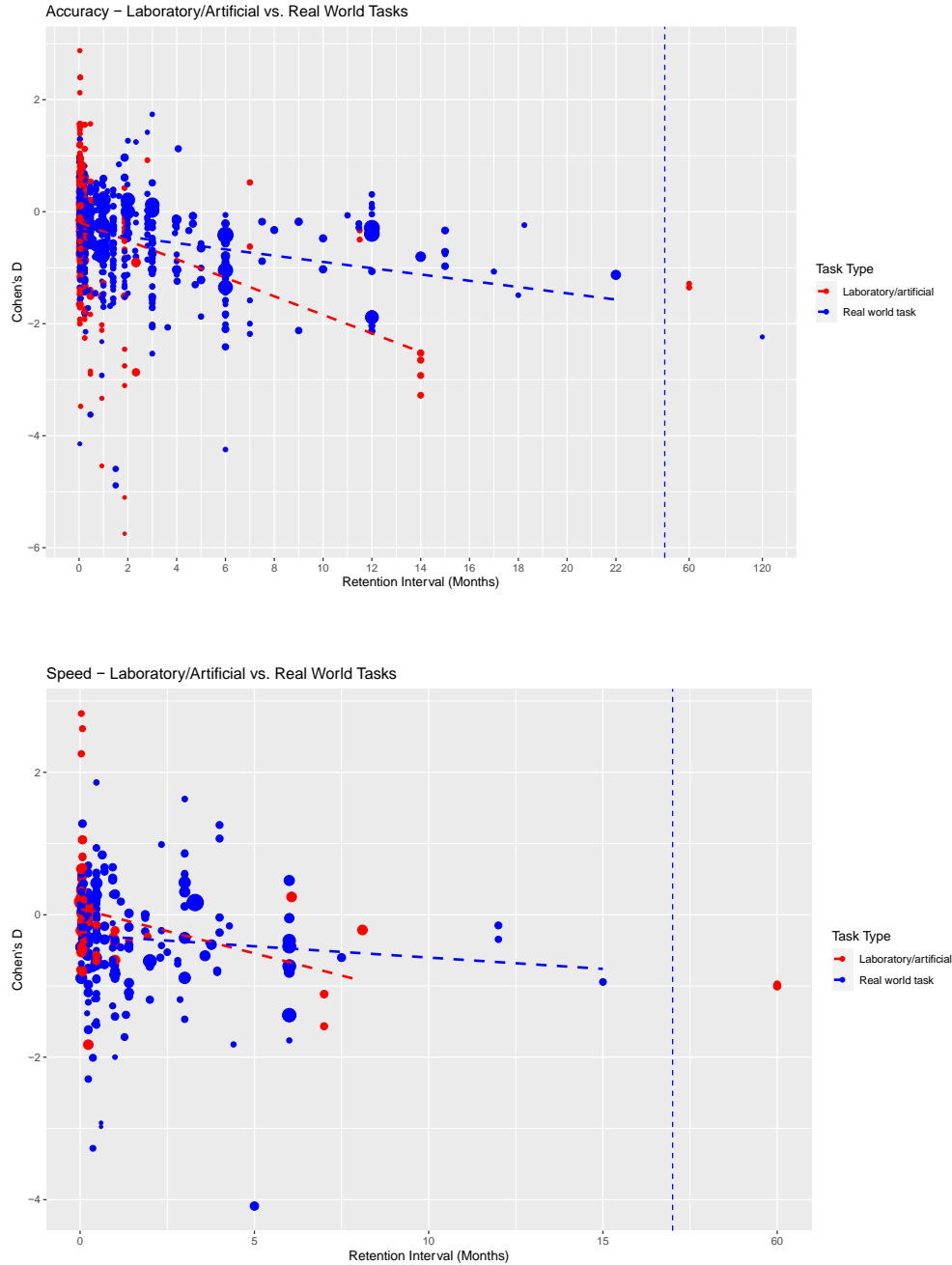


Figure 7 - Laboratory/Artificial vs. Real-World Tasks

Note. Scatterplot of effect sizes by retention interval for accuracy and speed-based performance measures color-coded by laboratory vs. artificial task. Red = laboratory/artificial task, blue = real-world task. Dotted lines indicate the linear relationship between retention interval and effect size for each task type. The two outlier retention intervals of 60 and 120 months are separated in these graphs by blue vertical-dotted lines and were not included in meta-regression analyses, but are included here to ensure all eligible data are reported.



Figure 8 - All Task Types

Note. Scatterplot of effect sizes by retention interval for accuracy and speed-based performance measures color-coded by task type. Red = laboratory/artificial task, blue = medical/dental task, green = military/transportation task, orange = sports task, purple = miscellaneous task. Dotted lines indicate the linear relationship between retention interval and effect size for each task type. The two outlier retention intervals of 60 and 120 months are separated in these graphs by blue vertical-

dotted lines and were not included in meta-regression analyses, but are included here to ensure all eligible data are reported.

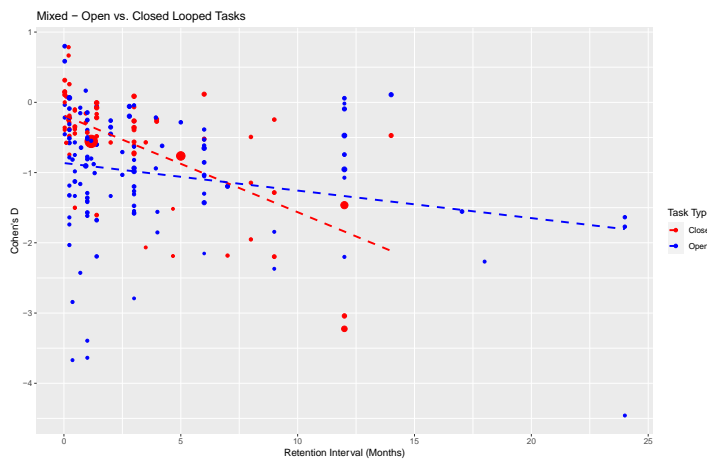
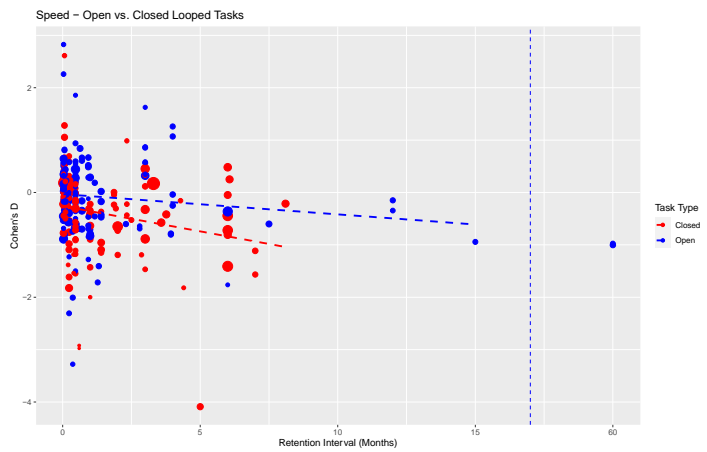


Figure 9 - Open vs. Closed-Looped Tasks

Note. Scatterplot of effect sizes by retention interval for accuracy, speed, and mixed performance measures color-coded by open vs. closed-loop task. Red = closed-looped, blue = open-looped.

Dotted lines indicate the linear relationship between retention interval and effect size for each task type. The two outlier retention intervals of 60 and 120 months are separated in these graphs by blue vertical-dotted lines and were not included in meta-regression analyses, but are included here to ensure all eligible data are reported.

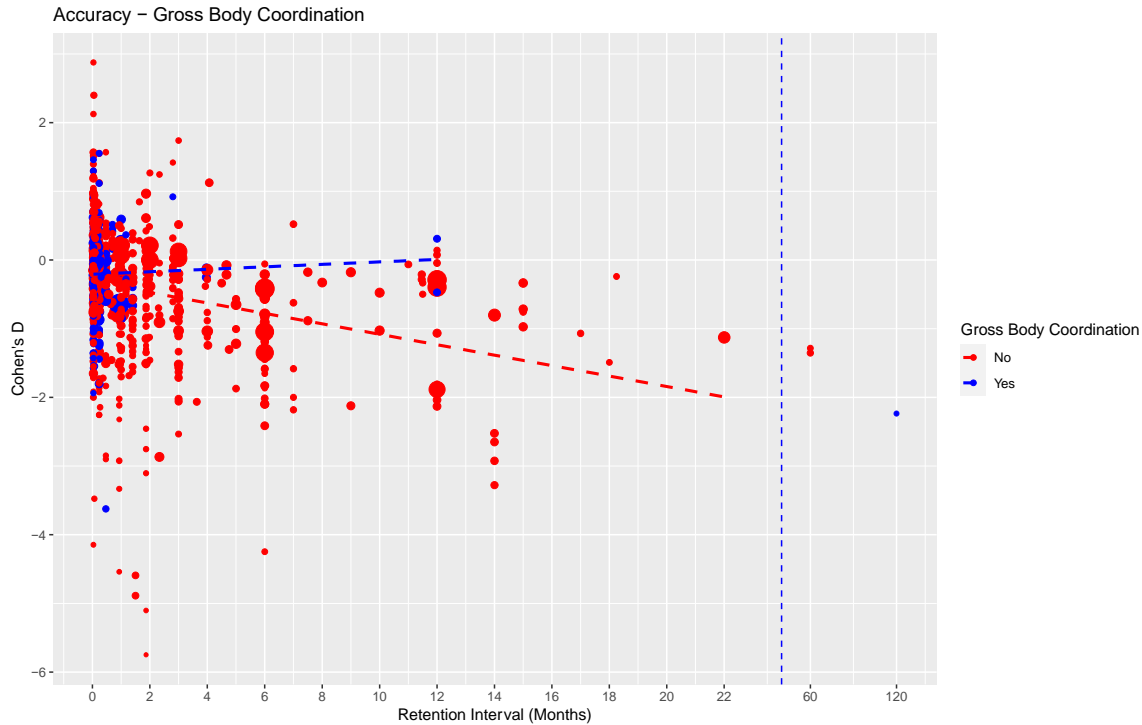


Figure 10 – Gross Body Coordination for Accuracy Performance

Note. Scatterplot of effect sizes by retention interval for accuracy-based performance measures color-coded by gross body coordination requirements. Red = no gross body coordination requirements, blue = gross body coordination requirements. Dotted lines indicate the linear relationship between retention interval and effect size for each task type. The two outlier retention intervals of 60 and 120 months are separated in these graphs by blue vertical-dotted lines and were not included in meta-regression analyses, but are included here to ensure all eligible data are reported.

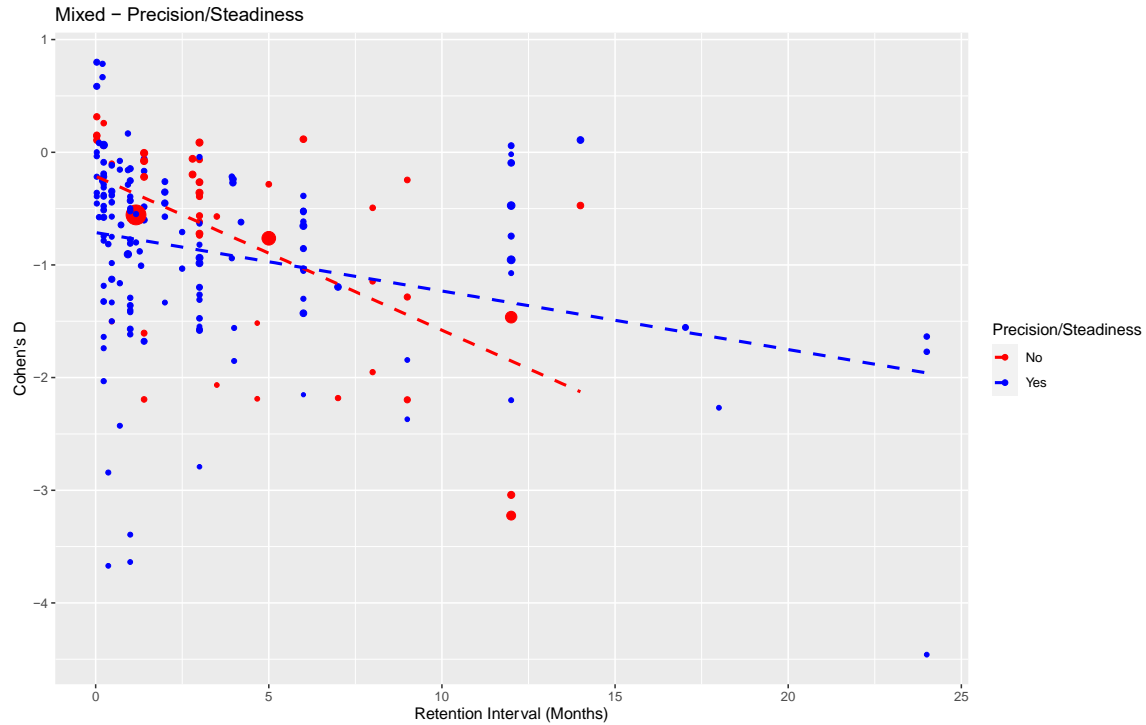


Figure 11 – Precision/Steadiness for Mixed Performance

Note. Scatterplot of effect sizes by retention interval for mixed performance measures color-coded by precision/steadiness requirements. Red = no precision/steadiness requirements, blue = precision/steadiness requirements. Dotted lines indicate the linear relationship between retention interval and effect size for each task type. The two outlier retention intervals of 60 and 120 months are separated in these graphs by blue vertical-dotted lines and were not included in meta-regression analyses, but are included here to ensure all eligible data are reported.

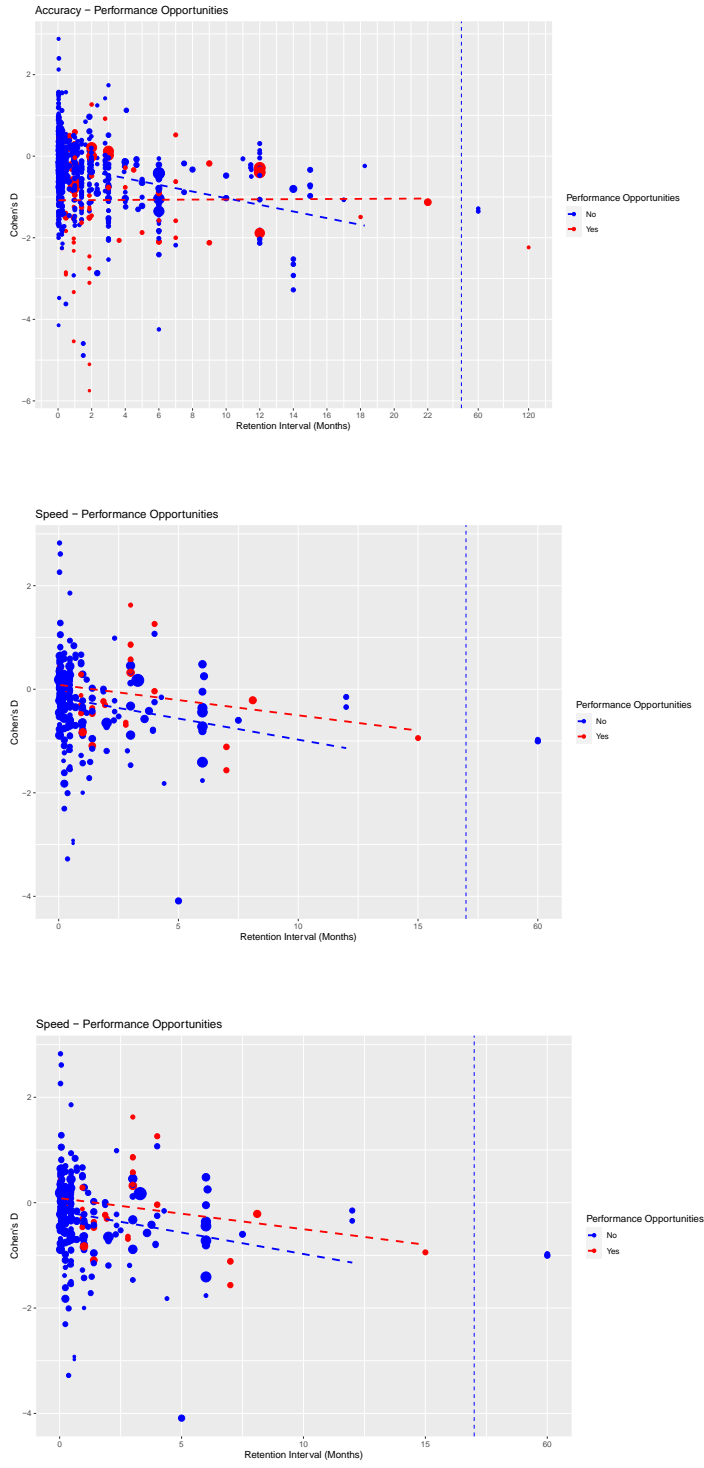


Figure 12 - Infrequent Performance Opportunities During Retention Interval

Note. Scatterplot of effect sizes by retention interval for accuracy, speed, and mixed performance measures color-coded by performance opportunities during retention interval. Red = performance opportunities during retention interval, blue = no performance opportunities during retention

interval. Dotted lines indicate the linear relationship between retention interval and effect size for each task type. The two outlier retention intervals of 60 and 120 months are separated in these graphs by blue vertical-dotted lines and were not included in meta-regression analyses, but are included here to ensure all eligible data are reported.

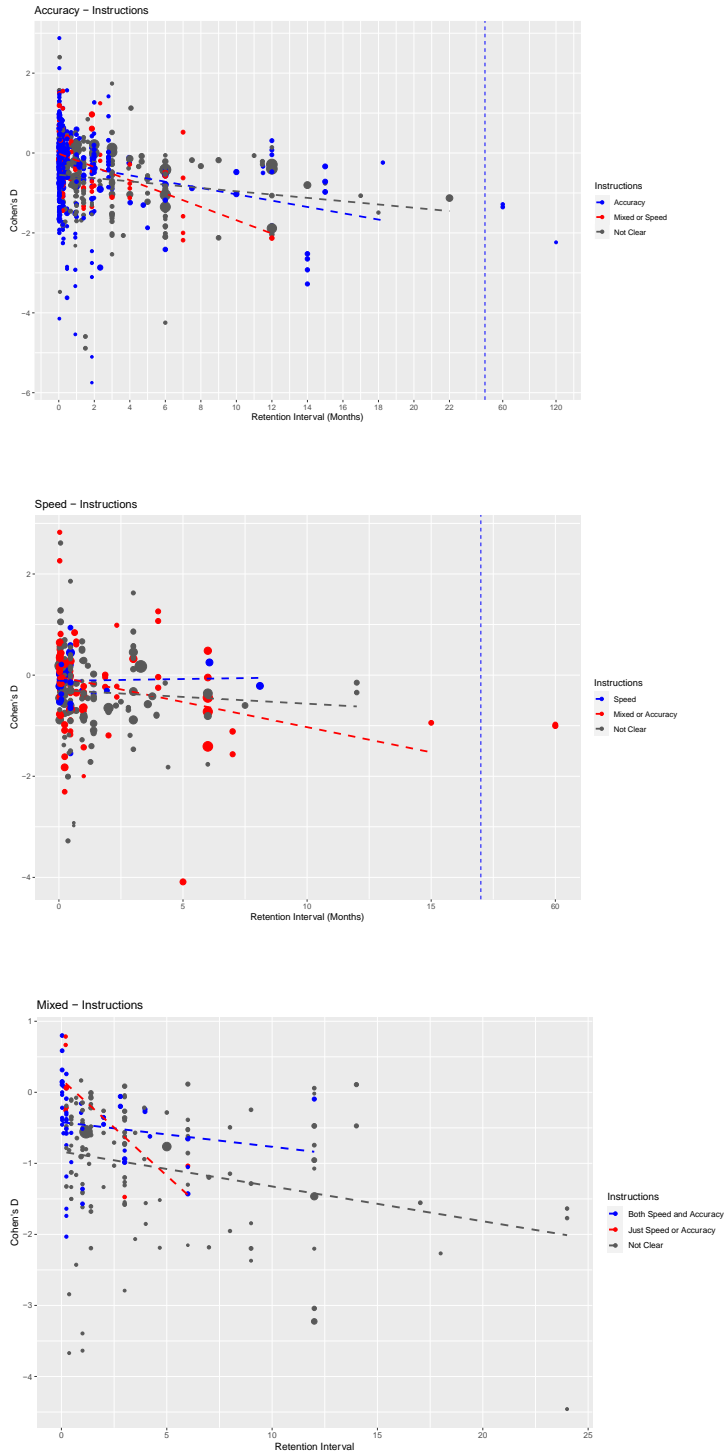


Figure 13 – Task Instructions

Note. Scatterplot of effect sizes by retention interval for accuracy, speed, and mixed performance measures color-coded by task instructions. Red = instructions aligned with performance measure, blue instructions not aligned with performance measure, gray = instructions not clear. Dotted lines

indicate the linear relationship between retention interval and effect size for each task type. The two outlier retention intervals of 60 and 120 months are separated in these graphs by blue vertical-dotted lines and were not included in meta-regression analyses, but are included here to ensure all eligible data are reported.

Table 11 – Narrative Summary of Moderator Analyses

Moderator	Effect	Summary of Results
Laboratory/artificial vs. real-world task	<p>Accuracy: Laboratory ↓ Real-world*</p> <p>Speed: Laboratory ↓ Real-world</p>	<p>It appears that laboratory/artificial vs. real-world task is a meaningful moderator of the relationship between retention interval length and skill decline, such that performance is better retained over time on real-world tasks than it is on laboratory/artificial tasks. The lack of significant results for speed was likely a result of low power.</p>
Broad task type	Accuracy: Laboratory ↓ Medical/dental*	<p>The moderating effect of task type is based primarily on differences between laboratory/artificial tasks and real-world tasks. Performance on laboratory tasks declines faster than performance on all real-world tasks, at least for accuracy. There is not sufficient evidence that there are meaningful differences in skill decay patterns between medical/dental, military/transportation, sports, and miscellaneous task.</p>
	Accuracy: Laboratory ↓ Military/transportation*	
	Accuracy: Laboratory ↓ Sports*	
	Accuracy: Laboratory ↓ Miscellaneous*	
	Speed: Laboratory ↓ Medical/dental	
	Speed: Military/transportation ↓ Laboratory	
Speed: Sports ↓ Laboratory		

	Speed: Laboratory ↓ Miscellaneous*	
Open vs. closed-looped task	Accuracy: Closed-looped ≈ Open-looped	There are mixed results for whether the moderating effect of open vs. closed-looped tasks. It appeared to be a meaningful moderator for mixed and speed performance measures, but not for accuracy performance measures.
	Speed: Closed-looped ↓ Open-looped	
	Mixed: Closed-looped ↓ Open-looped*	
Task complexity	Accuracy: Low component ↓ High component*	It appears that task complexity is a meaningful moderator of the relationship between retention interval and performance as the overall pattern of results suggests that performance on high complexity tasks is better retained than performance on low complexity tasks is. Evidence for the moderating effect is particularly strong for mixed performance measures and component complexity.
	Accuracy: Low coordinative ≈ High coordinative	
	Accuracy: Low dynamic ≈ High dynamic	
	Speed: Low component ≈ High component	
	Speed: Low coordinative ≈ High coordinative	
	Speed: Low dynamic ↓ High dynamic	
	Mixed: Low component ↓ High component*	
	Mixed: Low coordinative ↓ High coordinative*	
Mixed: Low dynamic ↓ High dynamic*		

Task ability demands

Accuracy: No dexterity ↓ Dexterity

Accuracy: Gross body ↓ No gross body*

Accuracy: Precision ≈ No precision

Speed: Dexterity ↓ No dexterity

Speed: Gross body ≈ No gross body

Speed: Precision ≈ No precision

Mixed: Dexterity ↓ No dexterity*

Mixed: Gross body ≈ No gross body

Mixed: No precision ↓ Precision*

Performance on tasks that require gross body coordination (for accuracy-based performance measures) and precision/steadiness (for mixed performance measures) may be better retained over time, but the high and low base rates for task ability demands in the current dataset make it difficult to draw conclusions regarding the moderating effect of task ability demands.

Performance opportunities

Accuracy: No perf. opp. ↓ Perf. opp*

Speed: Perf. opp. ≈ No perf. opp

Mixed: No perf. opp. ↓ Perf. opp.

It appears that practice opportunities are a meaningful moderator of the relationship between retention interval and skill decline, at least for accuracy and mixed performance measures. Performance is better retained over time when infrequent performance opportunities are available during the retention interval. The results do not extend to speed-based performance.

	Accuracy: Not Aligned ↓ Aligned*	The overall pattern of results suggests that task instructions do meaningfully moderate the relationship between retention interval and skill decline, such that performance is better retained when instructions are aligned with the performance measures.
Task instructions	Speed: Not aligned ↓ Aligned	
	Mixed: Not aligned ↓ Aligned	

Note. ↓ = more rapid skill decline than, ≈ = little or no difference in skill decline patterns, * = statistically significant. perf. opp. = performance opportunities during retention interval, gross body = gross body coordination, precision = precision/steadiness.

5.4 RQ 9: Loss of Skill Acquisition Gains

Thus far, I have reported results related to several dimensions of research on procedural skill retention. First, I began with a replication of the Arthur et al. (1998) skill retention meta-analysis using inclusion criteria of the current study to determine how the results would compare to their original findings. The results corroborated a trend of increasing decay over time, particularly at retention intervals longer than six months, but revealed lower levels of decay at shorter retention intervals. Second, meta-regression was used to quantify the level of procedural skill decay that can be expected per month for accuracy (a decline of 0.08 standard deviation units), speed (a decline of 0.06 standard deviation units), and mixed (a decline of 0.06 standard deviation units) performance. Third, a variety of task-related and study design moderators were explored, and laboratory/artificial vs. real-world task, task complexity, performance opportunities, and task instructions were identified as potentially meaningful moderators of skill decline rates. I now turn my attention to a vital component of the scope of this research that I have not yet addressed: Skill acquisition. The analysis presented in this section is designed to answer the fundamental applied question of how much of initial skill acquisition gains are retained over time.

To this point, skill retention has been quantified using meta-analytic methods. In order to answer the question of interest in this section, it was necessary to quantify skill acquisition using meta-analytic methods as well. Accordingly, a random-effects meta-analysis was computed using the effect sizes that capture differences in performance between initial acquisition and final acquisition. The primary difference between the two meta-analytic estimates was the absence of time, which only applied to research questions related to retention. A random effects meta-analysis rather than a meta-regression equation was computed because of the absence of the continuous

variable of retention interval in the analysis. Each performance measure type (i.e., accuracy, speed, mixed) was analyzed as a subgroup in the meta-analysis.

In total, 793 effect sizes were included in the skill acquisition meta-analysis (560 accuracy-based performance measures, 188 speed-based performance measures, and 115 mixed performance measures). The results of the meta-analysis are presented in Table 12. The meta-analytic estimate for the skill acquisition effect size for accuracy-based performance measures was $\delta = 1.49$ [1.38, 1.59], for speed-based performance measures was $\delta = 1.89$ [1.67, 2.10], and for mixed performance measures was $\delta = 2.47$ [2.17, 2.77]. The lack of overlapping confidence intervals between the three estimates indicates that these estimates were significantly different from one another. That is, skill acquisition gains for mixed performance measures were significantly larger than skill acquisition gains for speed-based performance measures, which were in turn significantly larger than skill acquisition gains for accuracy-based performance measures.

Table 12 – Skill Acquisition Meta-Analysis

	k	δ	SE	<i>t</i>	95% CI	τ^2	<i>I</i> ²
Accuracy	560	1.49	0.06	26.78**	[1.38, 1.59]	1.02	88.44%
Speed	188	1.89	0.11	17.33**	[1.67, 2.10]	1.55	89.38%
Mixed	115	2.47	0.15	16.18**	[2.17, 2.77]	1.95	92.47%

Note. k = number of effect sizes, SE = standard error, CI = confidence interval.

In order to determine how much of initial skill acquisition gains are lost over time with infrequent use, the intercept and the absolute value of the predicted Cohen’s *D* at various retention intervals yielded by the skill retention meta-regression were subtracted from the effect size yielded by the skill acquisition meta-analysis (see Table 13). This made it possible to estimate the percentage of initial skill acquisition gains that could be expected at different retention interval

lengths and the amount of time it would take to lose all initial gains if skills are not used (i.e., when the difference between initial acquisition gains and the absolute value of predicted retention Cohen's *D* approached zero). The results of this analysis are plotted in Figure 14.

The results suggest that skill acquisition gains for accuracy-based performance are lost more rapidly than skill acquisition gains for speed and mixed performance. For accuracy-based performance measures, one can expect about half of initial skill acquisition gains to be lost after between 6 and 9 months of infrequent use and all of initial skill acquisition gains to be lost just after approximately 15 months of infrequent use. For speed-based performance measures, one can expect about half of initial skill acquisition gains to be lost after between 12 and 15 months of infrequent use and all of initial skill acquisition gains to be lost just after approximately 27 months of non-use. For mixed performance measures, one can expect about half of initial skill acquisition gains to be lost after between 9 and 12 months of infrequent use and all of initial skill acquisition gains to be lost just after approximately 30 months of infrequent use. It is worth noting that mixed performance measures consisted largely of medical/dental tasks and as a result, the acquisition statistics for mixed performance may not be directly comparable to the accuracy or speed-based performance measures because of substantial differences in the tasks associated with each category. These results provide a clear answer to Research Question 9 based on the data available in the skill retention literature.

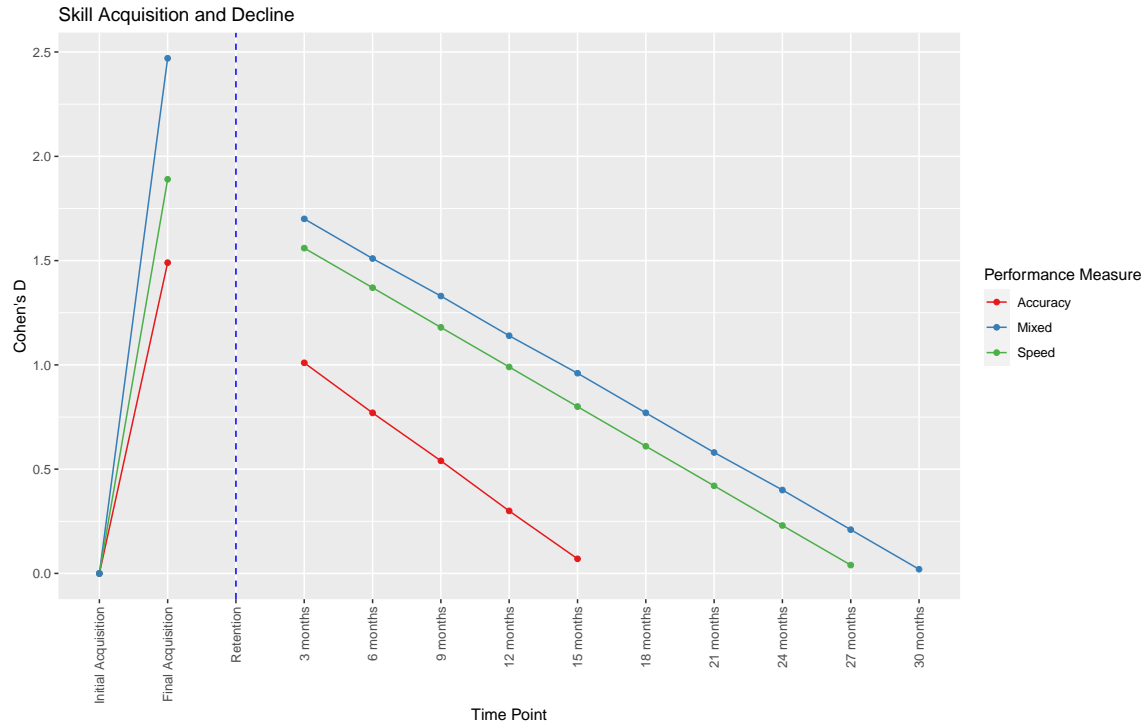


Figure 14 - Skill Retention in the Context of Initial Skill Acquisition Gains

Note. This figure shows the rise and decline of skills through stages of acquisition and retention. Acquisition and retention stages are separated by the blue dashed-vertical line. The effect sizes on the Y-axis are meant to represent the difference between skill levels prior to acquisition and skill levels at a given point in time.

Table 13 – Skill Retention in the Context of Initial Acquisition Gains

Time Point	Accuracy		Speed		Mixed	
	Cohen's <i>D</i>	% of Acquisition	Cohen's <i>D</i>	% of Acquisition	Cohen's <i>D</i>	% of Acquisition
Initial Acquisition	0.00	0.00	0.00	0.00	0.00	0.00
Final Acquisition	1.49	100.00	1.89	100.00	2.47	100.00
3 months	1.01	67.79	1.56	82.54	1.70	68.83
6 months	0.77	51.68	1.37	72.49	1.51	61.13
9 months	0.54	36.24	1.18	62.43	1.33	53.85
12 months	0.30	20.13	0.99	52.38	1.14	46.15
15 months	0.07	4.70	0.80	42.33	0.96	38.87
18 months			0.61	32.38	0.77	31.17
21 months			0.42	22.22	0.58	23.48
24 months			0.23	12.17	0.40	16.19
27 months			0.04	2.12	0.21	8.50
30 months					0.02	0.81

Note. This table shows the rise and decline of skills through stages of acquisition and retention. Effect sizes indicate differences in performance prior to acquisition and at a given point in time.

5.5 RQ 10: Comparison of Effect Sizes Computed with Different Formulas

The results presented to this point included only the effect sizes that were computed using the effect size formula that is typically used for independent samples because this is the form in which most data were reported, the approach allowed for isolation of retention from relearning effects when possible, and the approach is consistent with previous meta-analyses of skill retention (Arthur et al., 1998; Wang et al., 2013). However, there was another set of effect sizes coded that were not computed using the independent samples effect size formula. These effect sizes were either reported by the source authors as Cohen's *D* statistics, or they were reported by authors as paired-samples *t*-values or repeated measures ANOVA main effects (i.e., *F*-values) that were then converted to Cohen's *D* statistics. These effect sizes were analyzed separately because the differences in how they were computed warranted separate treatment based on the results of the Monte Carlo simulations.

A total of 71 effect sizes (61 converted from paired-sample *t*-values and 10 reported as *D* statistics) across 33 studies was analyzed. Effect sizes were once again analyzed separately for each performance measure category. There were 39 effect sizes representing accuracy-based performance measures, 12 effect sizes representing speed-based performance measures, and 20 effect sizes representing mixed performance measures. Moderator analyses were not conducted with this pool of studies because of the modest number of sources that reported effect sizes using paired-samples formulas. However, the meta-regression approach from the previous section was replicated in order to compare the overall strength of the effect sizes to those computed by the independent samples formula and to compare trends over time between the two groups of effect sizes. Specifically, a

random-effects meta-regression equation was computed with retention interval predicting effect size.

Results of each meta-regression are presented in Table 14 (accuracy), Table 15 (Speed), and Table 16 (mixed). Results indicate that the relationship between retention interval and effect size strength is not substantially different between the two methods of effect size calculation for accuracy-based and mixed performance measures ($\beta_{\text{Independent, Accuracy}} = -0.08, [-0.10, -0.06]$ vs. $\beta_{\text{Paired, Accuracy}} = -0.07, [-0.16, 0.02]$; $\beta_{\text{Independent, Mixed}} = -0.06, [-0.09, -0.04]$ vs. $\beta_{\text{Paired, Mixed}} = -0.05, [-0.13, 0.02]$). While the results were not statistically significant, the difference between beta-coefficients for speed-based performance measures in the two different pools of effect sizes was noticeably larger ($\beta_{\text{Independent, Speed}} = -0.06, [-0.11, -0.02]$ vs. $\beta_{\text{Paired, Speed}} = -0.02, [-0.17, 0.13]$). However, given that there were only 12 effect sizes computed with paired samples formulas for speed-based performance measures, these results should be interpreted with caution.

There was, however, a meaningful difference between the meta-regressions for the two pools of effect sizes that should be noted. The beta-coefficients associated with the intercept for the meta-regressions of effect sizes computed with the independent samples formula were smaller (i.e., stronger in magnitude) than to the beta-coefficients associated with the intercept for the meta-regressions of effect sizes computed with paired samples formulas across all three categories of performance measures ($\beta_{\text{Independent, Accuracy}} = -0.25 [-0.31, -0.19]$ vs. $\beta_{\text{Paired, Accuracy}} = -0.07 [-0.28, 0.15]$; $\beta_{\text{Independent, Speed}} = -0.14 [-0.24, -0.03]$ vs. $\beta_{\text{Paired, Speed}} = 0.00, [-0.44, 0.44]$; $\beta_{\text{Independent, Mixed}} = -0.58, [-0.73, -0.44]$ vs. $\beta_{\text{Paired, Mixed}} = -0.06, [-0.38, 0.26]$). This suggests that the overall magnitude of the effect sizes computed with the independent samples formula was larger than the magnitude of the effect sizes

computed with paired samples formulas when the retention interval was zero. While this result does address Research Question 10, it is surprising because it conflicts with expectations based on the Monte Carlo simulation reported earlier and is explored further in the Discussion chapter.

Table 14 – Meta-Regression of Paired Samples Effect Sizes for Accuracy

Accuracy (k = 39)				
Model	β	<i>SE</i>	<i>t</i>	95% CI
Intercept	-0.07	0.11	-0.65	[-0.28, 0.15]
Retention Interval	-0.07	0.04	-1.67	[-0.16, 0.02]
Tests				
Heterogeneity		Moderators		
$Q(37) = 104.55^{**}$, $F(1, 37) = 2.77$				
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.15	0.38	71.72%	3.54	7.10%

Note. * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes. The test of residual heterogeneity is Cochran’s Q test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size.

Table 15 – Meta-Regression of Paired Samples Effect Sizes for Speed

Speed (k = 12)				
	β	<i>SE</i>	<i>t</i>	95% CI
Intercept	0.00	0.20	-0.02	[-0.44, 0.44]
Retention Interval	-0.02	0.07	-0.26	[-0.17, 0.13]
Tests				
Heterogeneity		Moderators		
$Q(10) = 32.73^{**}$, $F(1, 10) = 0.07$				
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.18	0.42	69.06%	3.23	0.00%

Note. * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes. The test of residual heterogeneity is Cochran’s Q test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The

test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size.

Table 16 – Meta-Regression of Paired Samples Effect Sizes for Mixed

Mixed (k = 20)				
Model	β	<i>SE</i>	<i>t</i>	95% CI
Intercept	-0.06	0.15	-0.40	[-0.38, 0.26]
Retention Interval	-0.05	0.04	-1.48	[-0.13, 0.02]
Tests				
Heterogeneity		Moderators		
$Q(18) = 47.93^{**}$, $F(1, 18) = 2.02$				
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.11	0.33	64.40%	2.81	0.00%

Note. * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes. The test of residual heterogeneity is Cochran’s Q test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size.

5.6 Publication Bias

As previously mentioned, it is critical in any meta-analysis to account for the extent to which publication bias may have contributed to the observed results, and publication bias was evaluated in the current study using a two-pronged approach. The publication bias assessment results associated with each performance measure are therefore presented in two parts. Results for meta-analyses/meta-regressions with literature type (published vs. grey literature) as a moderator are presented, followed by plots of effect size magnitude by precision and results for PET-PEESE tests. All results are reported for effect sizes representing each of the three performance measure types.

The inclusion of grey literature in the literature review strategy itself was designed to mitigate the likelihood of publication bias. However, to assess publication bias directly

in the studies included, literature type was tested as a moderator of effect size strength and assessed using sub-group analyses separately for accuracy-based, speed-based, and mixed performance measures using the studies included in the primary analyses (i.e., the effect sizes computed using independent samples, excluding the two studies with outlier retention intervals). Results are reported in Table 17. While the results suggest that differences in effect size strength are not statistically significant (confidence intervals for published vs. grey literature overlap for all three performance measure categories), interestingly, the results reveal a trend of stronger effect sizes for grey literature in comparison to published literature for accuracy-based and speed-based performance measures (Accuracy: $\delta_{\text{Published}} = -0.35 [-0.41, -0.29]$, $\delta_{\text{Grey}} = -0.53 [-0.71, -0.36]$; Speed: $\delta_{\text{Published}} = -0.19 [-0.29, -0.09]$, $\delta_{\text{Grey}} = -0.53 [-0.71, -0.36]$), but in the opposite direction for mixed performance measures (Mixed: $\delta_{\text{Published}} = -0.83 [-0.97, -0.70]$, $\delta_{\text{Grey}} = -0.58 [-0.81, -0.35]$). In the case of mixed performance measures, the differences may be attributed to the fact that the average retention interval for published studies was nearly two months longer than the average retention interval for grey literature ($M_{\text{Published}} = 3.78$, $M_{\text{Grey}} = 1.85$, $t(50.85) = 2.86$, $p < 0.01$). However, the same cannot be said of accuracy or speed-based performance measures as the average retention interval lengths for accuracy were similar ($M_{\text{Published}} = 1.52$, $M_{\text{Grey}} = 1.51$, $t(219.37) = 0.04$, *ns*) and average retention intervals for speed-based performance measures were longer for published literature than they were for grey literature ($M_{\text{Published}} = 1.32$, $M_{\text{Grey}} = 0.79$, $t(138.26) = 2.45$, $p < 0.05$), even though grey literature retention effect sizes were larger.

The influence of publication bias on estimates of skill decay over time in studies included in the meta-analysis was directly assessed by testing literature type as a moderator

in the meta-regression equations previously reported in which retention interval was used to predict effect size magnitude. Published literature was chosen as the reference group for publication type, meaning that the beta-coefficient for the interaction term represented a change in the slope of the regression line representing the relationship between retention interval and Cohen's D when a study was unpublished compared to when it was published. Literature type was found to be a significant moderator of the relationship between retention interval for accuracy-based performance measures ($\beta = 0.09$ [0.04, 0.15]), but not for speed-based ($\beta = -0.01$ [-0.21, 0.20]) or mixed performance measures ($\beta = 0.01$ [-0.12, 0.13]). The positive beta coefficient associated with the significant effect for accuracy suggests that skills were better retained over time in grey literature than they were in published literature. Results for accuracy-based performance measures are plotted in Figure 15.

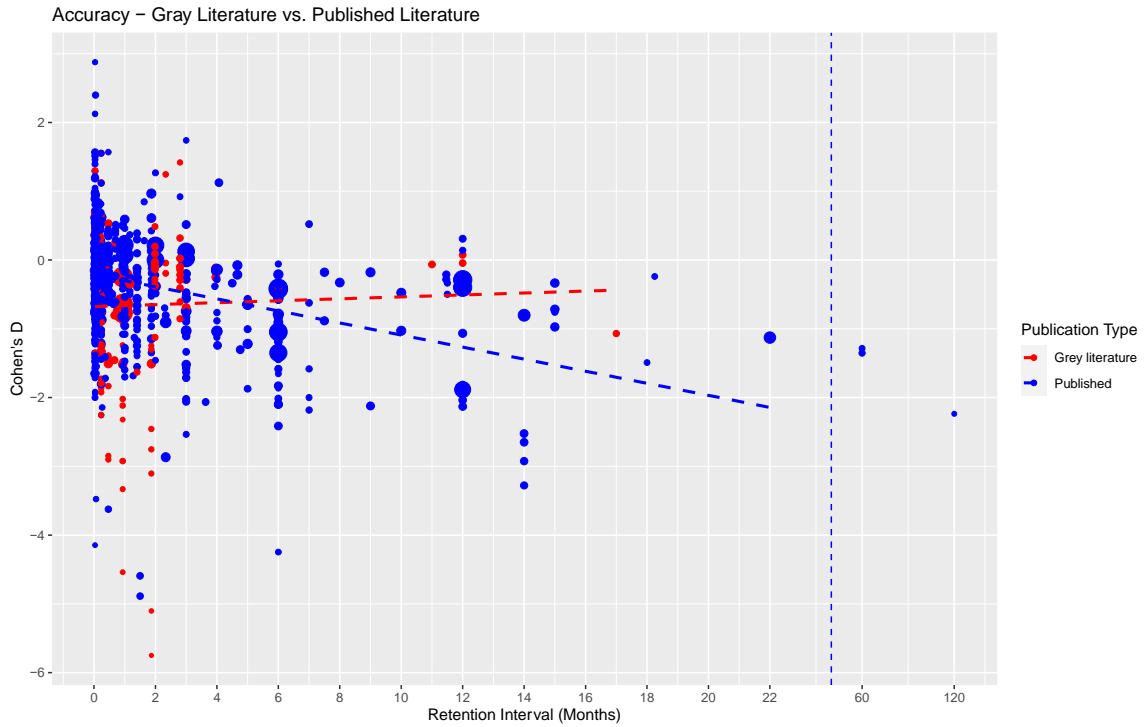


Figure 15 - Retention Trends for Accuracy Published vs. Grey Literature

Note. Scatterplot of effect sizes by retention interval for accuracy-based performance measures color-coded by literature type. Red = grey literature, blue = published literature. Dotted lines indicate the linear relationship between retention interval and effect size for each task type. The two outlier retention intervals of 60 and 120 months are separated in these graphs by blue vertical-dotted lines and were not included in meta-regression analyses, but are included here to ensure all eligible data are reported.

Table 17 – Moderating Effect of Publication Types on Effect Size Strength

	Accuracy		Speed		Mixed	
	Published	Grey	Published	Grey	Published	Grey
k	698	131	222	45	155	25
δ	-0.35	-0.53	-0.19	-0.34	-0.83	-0.58
95% CI	[-0.41, -0.29]	[-0.71, -0.36]	[-0.29, -0.09]	[-0.58, -0.10]	[-0.97, -0.70]	[-0.81, -0.35]
τ^2	0.47	0.62	0.37	0.37	0.57	0.08
I^2	84.96%	87.13%	73.56%	73.50%	89.81%	33.17%
H^2	6.65	7.77	3.78	3.77	9.82	1.50

Note. k = number of effect sizes, CI = confidence interval.

While these analyses speak to the direct presence of publication bias in the studies included in the meta-analysis, they do not speak to studies that may not have been included. The relationship between precision and effect-size magnitude was evaluated to determine the likelihood that publication bias may have prevented some results from being included. The relationship was assessed via PET-PEESE tests, which involve a meta-regression of the effect sizes predicted by an index of precision associated with each effect size (in this case, the square root of the retention sample size for PET and the retention sample size for PEESE tests). The results of the PET-PEESE tests are reported as regression coefficients and significance tests and are also included on plots of effect size strength by the square root of the sample size in Figure 16.

The plots suggest that there may in fact be some evidence of publication bias, as there are a number of data points in the bottom left-hand corner of both the accuracy and mixed performance measure plots that indicate that the particularly large effect sizes come from studies with small samples. However, PET-PEESE results suggest that on the whole, this potential bias may not be a large concern. The PET-PEESE tests for speed-based and mixed measures were not significant (PET_{Speed}: $t(265) = -1.94$, *ns*; PEESE_{Speed} = $t(265) = -1.89$, *ns*; PET_{Mixed}: $t(178) = 0.06$, *ns*; PEESE_{Mixed} = $t(178) = -0.14$, *ns*). PET results for accuracy were significant, but the results for the PEESE test were not (PET_{Accuracy}: $t(827) = -2.81$, $p < 0.05$; PEESE_{Accuracy}: $t(827) = -1.81$, *ns*). On one hand, the significant PET test along with the visible tail on the left side of the distribution of the plot of effect size by precision may indicate some publication bias. But, on the other hand, the non-significant PEESE test tends to give more accurate estimates at higher levels of heterogeneity and the accuracy meta-regression was characterized by substantial heterogeneity ($I^2 = 83.21\%$).

In sum, there is some evidence of publication bias indicated in these analyses. There was a non-trivial difference in skill decay effect-size magnitude between published and grey literature, literature type moderated the relationship between retention interval and skill decay for accuracy-based performance measures, and interpretation of plots of precision by effect size and PET-PEESE test results suggested the possibility of publication bias, particularly for accuracy-based performance measures. However, the degree of publication bias is not substantial, and it is unlikely to have substantially influenced the results for three primary reasons: First, while one would normally expect effect sizes to be stronger in published studies than in grey literature, to the extent that publication bias is indicated by the published and grey literature included, the opposite pattern is observed. Grey literature effect sizes were stronger than published literature effect sizes for accuracy and speed-based performance measures.

It may be the case that publishing bias worked in the opposite direction from what is typically expected, at least for accuracy and speed-based performance measures, because many of the effect sizes were from training studies for which reporting that skills are retained over time is more advantageous to publication than reporting that skills decline over time is. Therefore, while it is possible that there was a bias toward larger effect sizes at smaller sample sizes as the accuracy PET results and plot indicate, it is also possible that a bias toward small decay effect sizes in the training literature either balanced out this bias or even led to an underestimate of skill decay estimates. It is, however, worth noting that the more traditional publication bias trend was observed for the significant results for retention over time – effect sizes from published literature grew larger in magnitude over time than effect sizes from grey literature for accuracy-based performance measures.

Second, it may be that differences in observations between grey literature and published literature were a reflection of study quality. If, for example, grey literature consisted of more study-design flaws than published literature, then publication bias would have eliminated noise from the estimation of the true effect. Finally, while some publication bias likely exists to some degree in most meta-analyses, that bias is often relatively small and unlikely to confound reported results in most circumstances. Statistically significant results in the expected direction are only 1.17 times more likely to be published than results that are non-significant or significant in the opposite direction (Mathur & VanderWeele, 2021) and effect size magnitude and significance often do not differ strongly between published and unpublished literature or between journals of disparate quality (Dalton et al., 2012).

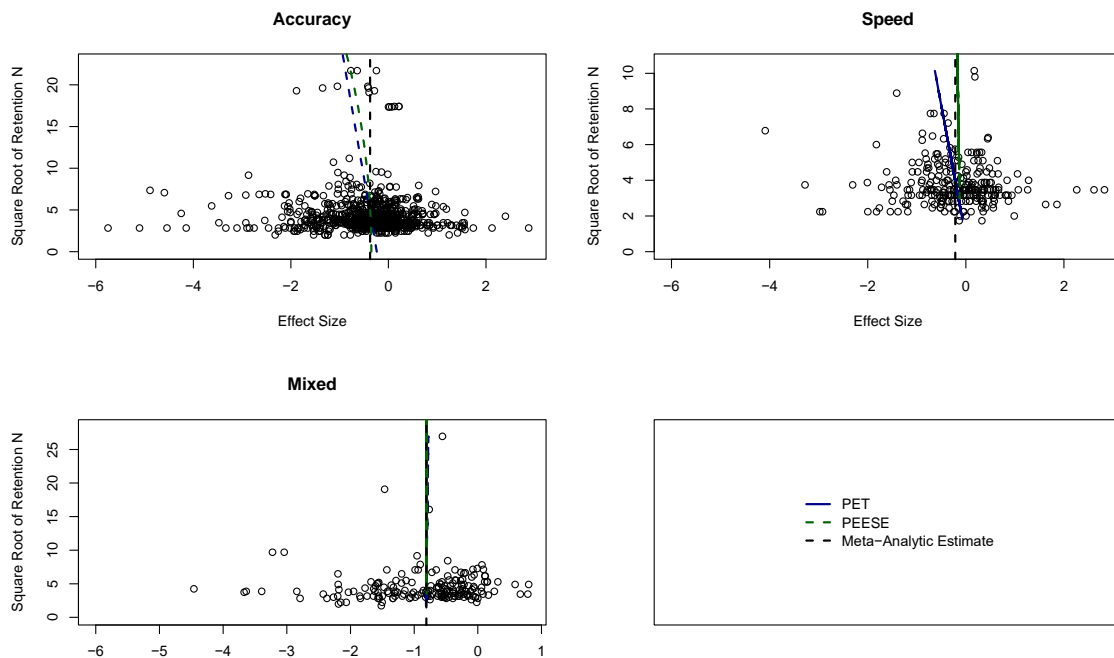


Figure 16 – Effect Size By Precision Plots

CHAPTER 6. DISCUSSION

I have now presented the findings of all analyses and discussed the extent to which they answer each of the 10 research questions that guided the current study. In the Discussion chapter, I provide a narrative review of each component of the results and its implications and propose explanations for what was found or not found. I step through each research question just as I did in prior chapters. However, several of the research questions are grouped together in this chapter because the results lend themselves to conceptually similar discussions. After stepping through the implications of the results for each research question, I conclude the chapter by highlighting limitations of the current study that lend themselves toward opportunities for future research.

6.1 Research Question 1: Replication of Arthur et al. (1998)

I began by replicating Arthur et al.'s (1998) meta-analytic approach using all eligible studies published 1998 or earlier in order to determine how their findings would have compared had they focused on procedural skills and searched the same databases. Ultimately, the replication yielded a similar overall conclusion that the original study did with respect to skill decay over time. That is, there is evidence of notable skill decay over time, particularly at retention intervals of six months or longer. There were, however, two notable differences worthy of discussion.

First, effect sizes were more modest at shorter retention intervals in the replication that focused only on procedural skills. In other words, procedural skills showed less decay within a week or two of final acquisition than Arthur et al.'s (1998) finding suggested. It

is possible that the discrepancy in effect size strength at shorter retention intervals largely reflects differences between the tasks that were included in each meta-analysis (i.e., procedural skills vs. all skills). For example, several studies included in the Arthur et al. (1998) meta-analysis involved the memorization of nonsense syllables. Given the artificial nature of such tasks, it is plausible that performance may decline more substantially over a shorter period of time than performance on a task with real-world relevance such as CPR, for example, would. These results are also consistent with Arthur et al.'s (1998) finding that performance on cognitive tasks decayed more than performance on physical tasks did. The difference between decay trends illustrates the difference between procedural skills and skills that do not require a significant motor component, and therefore the novel scope of this study. Given that procedural skills are at times needed in high-stakes situations with high degrees of time pressure (e.g., CPR, intubation, combat skills), the findings of the current study are particularly relevant to applied researchers or organizations interested in maintaining critical skills over intervals of infrequent use.

Second, there were 63 more effect sizes included in the replication than were included in the original study. Given the current study's more narrow focus on procedural skills rather than all skills, it may come as somewhat of a surprise that the current study included more data points within the same time period. However, there are several feasible explanations for this discrepancy. It may be that the current study's literature review yielded more relevant results because of a broader search including databases such as PubMed and Web of Science that were not included in Arthur et al. (1998) or because modern databases are more robust. It is also plausible that the decision to include multiple retention intervals or tasks from the same study led to a slight increase in the number of

available data points. Finally, it may be that I went to greater lengths to estimate effect sizes rather than exclude studies (e.g., estimation of standard deviations from error bars on graphs) than the original study authors did.

6.2 Research Questions 2 & 9: Skill Loss Estimates

The most fundamental questions that I sought to answer in this project involved quantitative estimates of skill. In Research Question 1, I asked about the level of skill decay that can be expected over time, and in Research Question 9, I considered these losses in the context of skill acquisition by asking what amount of skill acquisition gains would be lost over different intervals of time. Results indicated that: (1) On average, Cohen's *D* statistics declined by 0.08 SD units per month for accuracy-based performance measures and by 0.06 SD units per month for speed-based and mixed performance measures, and (2) on average, initial acquisition gains were lost for accuracy-based measures after 15 months of infrequent use, for speed-based performance measures after 27 months of infrequent use, and for mixed performance measures after 30 months of non-use, respectively.

The skill acquisition loss results should be taken with somewhat of a grain of salt because there were relatively large differences in initial acquisition gains across the three performance measures, which are likely related to the differences in tasks assessed within the three performance measure categories. For example, greater gains may have been observed for mixed performance measure tasks which were mostly medical/dental tasks because medical tasks were more difficult than many laboratory tasks. Participants may have had lower levels of initial performance prior to acquisition as a result. However, the results do suggest that skill proficiency levels gained during training are likely lost

somewhere between one year and two and half years after they are acquired, if they are not used consistently. These estimates can be used by those interested in maintaining skill proficiency to project how much skill decay can be expected within a particular period of time, and therefore when refresher training should be required or recommended.

The estimates of skill decay indicate substantial loss over longer periods of time (i.e., mean skill declines of a standard deviation unit after nine to 12 months for accuracy-based skills, 12 to 15 months for speed-based skills, and six to nine months for skills that require both speed and accuracy). However, based on some of the individual effect sizes observed, it is possible that these may even be underestimates. Surprisingly, 378 of the 1,352 retention effect sizes coded (including those calculated by the independent samples formula, converted from paired *t*-values, and reported by authors) were *greater than zero*, indicating that performance *increased* from final acquisition to retention. Many of the positive retention effect sizes may be explained by relatively short retention intervals where the retention test actually represented further acquisition rather than retention – 256 of the effect sizes came from studies with a retention interval of a week or less. However, this still leaves 122 retention effect sizes greater than zero from studies with retention intervals longer than one week. One generally would not expect performance of skills to increase after a week or more of non-use. The existence of positive effect sizes suggests that even if effect sizes indicated lower skill levels at retention in comparison to acquisition (i.e., negative effect sizes), it is possible the differences were underestimated. From my perspective, there are three potential explanations for the underestimation of skill retention effect sizes, each of which is described below.

First, skill retention effect sizes may have been underestimated because retention was confounded with relearning effects. The earliest studies of skill retention found that although motor skills decay over intervals of non-use, it takes substantially fewer trials to reach prior performance levels than it did initially (Hill, 1934, 1957; Hill et al., 1913; Swift, 1905, 1906, 1910). These findings suggest that the colloquial phrase “skills come back like riding a bicycle” more accurately captures the essence of relearning rather than retention. If an adult were to attempt riding a bicycle after years without doing so, they may wobble at first, but they will be able to ride the bike smoothly in a relatively short period of time. For a task where performance is relatively low stakes such as riding a bicycle, the initial wobbling is unlikely to cause adverse consequences. But, the consequences may be more severe if an individual needs to perform CPR and there is an early performance decrement. The quality of early compressions and ventilations may determine whether or not the recipient survives. It is therefore critical that skill retention studies isolate retention performance from relearning as much as possible.

Retention tests often consist of multiple trials or multiple blocks. For example, a retention test may consist of putting a golf ball 10 times or of three blocks of 10 putts. I attempted to isolate retention performance from relearning by only coding the first trial or first block for which adequate retention data were reported. However, this was not always possible. Authors often reported a single retention test score or scores for blocks rather than trials. If, for example, an individual missed their first putt of the retention test by over two feet but each of their remaining putts landed within six inches of the hole and the authors only reported a single retention test score or retention scores by block, superior performance on the later trials in which relearning occurred would confound the measure

of performance on the first putt. It may not always be possible to isolate retention from relearning effects, which presents a significant challenge for skill retention researchers. For example, many tasks may allow for relearning even within a single trial (e.g., open-looped tasks). But, relearning effects within a trial, within a block, or across blocks of a retention test likely led to an underestimate of the extent to which skills decline between acquisition and retention.

A second explanation for the possible underestimation of skill retention estimates is that estimates were confounded by speed-accuracy-tradeoff decisions. A notable challenge for skill retention researchers is making sense of two performance measures that are both practically important but are unlikely to be theoretically comparable – accuracy and speed. Skill retention researchers have three options when it comes to reporting performance data: (1) Aggregate speed and accuracy into a single composite measure of performance, (2) report only speed or accuracy data and risk neglecting some component of performance, and (3) report performance data separately for speed and accuracy. All three choices were represented in the literature. The first option, which the authors of 180 studies in the meta-analysis chose, is inherently problematic because increased accuracy often comes at the expense of decreased speed. There are some circumstances where the second option is plausible if it can be justified that one component of performance is far more crucial than the other, but in many circumstances performing a skill quickly and accurately can provide substantial benefits. However, given that speed and accuracy are not positively correlated at the individual level, it can be difficult to interpret changes in speed or accuracy over time. For example, perhaps an individual progressed faster through a task during retention than they did during acquisition, but made more errors. Their speed-

based performance measure would indicate improvement from final acquisition to retention, even though that does not reflect the reality of how skills have changed over time. Interpretation of these results is difficult because it is not possible to determine whether skill levels have declined, individuals have altered their speed-accuracy-trade-off decisions, or both.

Of the studies included in the meta-analysis, 69 reported effect sizes for both speed and accuracy-based performance measures, 41 of which involved retention intervals longer than one week. Of the 122 retention effect sizes greater than zero (i.e., skill improvement was observed after the retention period), 16 of them came from these 41 studies and were accompanied by a negative retention effect size for the opposite performance measure type. In other words, more accurate performance was accompanied by slower speed, or vice versa. While 16 effect sizes represent a small number of the overall results, these instances do illustrate the challenges associated with making sense of speed and accuracy simultaneously. It is also possible that some skill retention effect sizes were underestimated because greater performance decrements occurred for other performance measures (e.g., accuracy rather than speed) even if the effect sizes were not positive. The correlation between accuracy and speed-based effect sizes for all studies that reported both accuracy and speed-based performance measures was only $r = 0.16$ (a relatively small correlation), which further indicates that different performance measures may not have been equivalently reflective of skill levels. While more careful attention to these issues (e.g., analyzing speed and accuracy separately) may lead to more precise measurement of performance decrements, the fact that speed and accuracy are not conceptually similar but

both represent meaningful aspects of performance will likely remain a challenge in this area of research.

A final explanation for the potential underestimation of retention effect sizes is the possibility of performance opportunities or task exposure during the retention interval. Moderator analyses indicated that infrequent performance opportunities were a significant moderator of the relationship between retention interval and skill decay, such that skills were better retained when infrequent performance opportunities were available. Similar trends were observed for mixed performance measures (though they were not statistically significant). These results suggest that infrequent performance opportunities (either earlier retention tests or exposure to practice opportunities such as access to simulators, internships/rotations, etc.) may go a long way toward maintaining accuracy-based performance even if skills are not frequently used. It is quite possible that had researchers chosen to use separate samples in studies with multiple retention intervals (e.g., one sample tested for retention one month after training, another sample tested for retention six months after training), skill decay effect sizes would have been stronger. Additionally, while I attempted to code any circumstance in which performance opportunities were possible, it is highly likely that some authors did not report the possibility of exposure to performance opportunities during the retention interval.

6.3 Research Questions 3-8: Moderators of Skill Loss Trends

I also sought to explore both task-related and study-design moderators of the relationship between retention interval and skill decline. In these analyses, I investigated whether moderators previously identified in the skill retention literature applied

specifically to procedural skills and considered a broader set of moderators than previously investigated. Prior meta-analyses of skill retention assessed moderators through subgroup analyses that allowed for the determination of whether effect sizes differed across levels of moderators regardless of retention interval. However, my approach went a step further by considering whether trends of skill decline differ over time between levels of moderators. The results from moderator analyses were far from definitive, as many of the trends lacked consistency across all performance measures. Even when moderating trends were observed, results often lacked statistical significance for performance measures other than accuracy likely due to insufficient power resulting from fewer studies with longer retention intervals. Yet, the overall pattern of results suggests that the following variables are potentially meaningful moderators: (1) Task type, (2) task complexity, (3) performance opportunities during the retention interval, and (4) task instructions.

Based on the results, it appears that whether a task is a laboratory/artificial task or a real-world task is a meaningful moderator of the relationship between retention interval and skill retention. Performance on real-world tasks was better retained over time than performance on laboratory/artificial tasks for accuracy-based performance measures. Although the results were not statistically significant, likely due to a lack of power, the same trend was observed for speed-based performance measures. This extends Arthur et al.'s (1998) finding that performance on artificial tasks displays more decay generally than performance on natural tasks does by providing evidence that performance on artificial tasks also decays faster over time than performance on natural tasks does. The finding also clarifies uncertainty based on the small magnitude of differences reported in the prior meta-analysis. A plausible explanation may be that real-world tasks are more representative of

the environments that individuals are likely to practice and use their skills outside of an evaluation setting and participant motivation to perform well is therefore higher.

Results related to task complexity were also consistent with prior research on skill retention. Component complexity was a significant moderator of the relationship between retention interval and skill decay for accuracy-based performance measures and all three complexity dimensions (component, coordinative, and dynamic) were significant moderators of the relationship for mixed performance measures. In all complexity analyses, results suggested a general trend toward better retention over time for performance on high-complexity tasks than for low-complexity tasks. These results are consistent with Wang et al.'s (2013) findings that performance declined less for high-complexity tasks than for low-complexity tasks. A possible explanation for this trend may be that high-complexity tasks require more attention and effort to reach proficiency during acquisition and the investment of attention and effort leads to better retention. However, task type and complexity are inherently related. Real-world tasks had significantly higher task complexity ratings across all three dimensions than laboratory/artificial tasks did (Component: $M_{\text{Real}} = 1.78$, $M_{\text{Lab}} = 1.10$, $t(83.23) = 6.45$, $p < 0.01$; Coordinative: $M_{\text{Real}} = 2.29$, $M_{\text{Lab}} = 1.50$, $t(43.76) = 5.28$, $p < 0.01$; Dynamic: $M_{\text{Real}} = 1.75$, $M_{\text{Lab}} = 1.21$, $t(52.31) = 3.63$, $p < 0.01$). It may be that laboratory/artificial tasks are shortened for logistic reasons and are therefore less complex and not as well retained, or it may be that low complexity tasks are generally artificial and therefore not as well retained. The direction of the relationship remains unclear.

Performance opportunities were not investigated in prior research, but it appears that they may influence skill decay trends. When performance opportunities were available during the retention interval, either through earlier retention tests or task exposure,

accuracy-based performance was better retained over time than when performance opportunities were not available or not reported. The same trend was observed for mixed performance measures, although it was not statistically significant. These results are consistent with Wang et al.'s (2013) predictions, although the authors did not have sufficient data to test the relationship. Implications may be significant for organizations interested in avoiding employee skill decay, as there may be substantial benefits to having their employees engage in occasional retention tests. The benefits of retention tests seem to stretch beyond identifying at-risk individuals, as less rapid skill decline was observed on second retention tests, at least for accuracy-based and mixed performance measures. It was, however, surprising to find that skill decay patterns were similar for speed-based performance measures when performance opportunities were and were not available during the retention interval. It is possible that infrequent opportunities do more to maintain accuracy-based performance than they do speed, but there is no theoretical explanation or corroborating empirical evidence to support this. It may be that there weren't enough speed-based studies with performance opportunities to detect an effect or that authors who reported speed-based performance measures did not sufficiently report the possibility of performance opportunities.

The significant moderating effect of task instructions might not necessarily underscore genuine differences in skill decay trends. Instead, it could primarily illustrate how study design can influence speed-accuracy tradeoff decisions. Trends for all three performance measures were in the direction of better skill retention over time when task instructions aligned with the performance measure (i.e., accuracy-based instructions for accuracy-based performance measures, speed-based instructions for speed-based

performance measures, and both speed and accuracy-based instructions for mixed performance measures). In other words, the differences observed with respect to performance at final acquisition and retention became even more disparate over time when instructions were not aligned with how performance was measured. It is somewhat unlikely that the more rapid decay of skills for tasks with instructions that did not align with performance measures reflects true differences in skill levels because one would not expect instructions themselves to influence a participant's skills. It may be that differences in performance that grow over time are a result of differences in calculations that individuals are making with respect to speed-accuracy tradeoffs at acquisition vs. retention (e.g., being more careful, but proceeding slower, on a real-world task if one suspects a performance decay).

The results for the following moderators were, unfortunately, less illuminating: Open vs. closed-loop tasks, task ability demands, and real-world task category. Results for open vs. closed-looped tasks were inconsistent across performance measures and difficult to make sense of. Performance on open-looped tasks was significantly better retained than performance on closed-looped tasks only for mixed performance measures, and not for accuracy or speed. Trends indicated that performance on open-looped may have been slightly better retained for speed-based measures as well, but there was no difference in skill retention trends between open and closed-looped tasks for accuracy-based tasks. This is particularly perplexing because many tasks in classic skill acquisition and retention research that would be classified as open-looped are accuracy-based tasks (e.g., tracking tasks, pursuit rotor).

Given that most effect sizes for mixed performance measures were from medical/dental tasks, it may be that there is something about medical/dental tasks that allows open vs. closed-looped tasks to play a larger role in determining skill retention trends. However, it is more likely that these results in combination with contradictory results related to open vs. closed-looped tasks in the two primary meta-analyses on skill retention suggest limitations in the extent to which the classification of open vs. closed-looped tasks accounts for skill retention. The basic laboratory tasks used in early skill retention studies could be relatively easily classified as open or closed-looped. For example, it is relatively straightforward to classify a tracking task as an open-looped task or a finger-tapping task as a closed-looped task. However, many occupational skills consist of both open and closed-looped components. For example, some steps of performing a surgical procedure may require a surgeon to continuously monitor changing conditions and make adjustments accordingly (indicative of an open-looped task) while some steps may be carried out quickly and efficiently with a discrete beginning and end (indicative of a closed-loop task). It may also be difficult to isolate retention from relearning for open-looped tasks because open-looped tasks require continuous adjustment that may result in relearning even within a single trial. For these reasons, despite its prominence in early skill retention literature, the distinction between open vs. closed-loop tasks may not be as meaningful of a moderator of procedural skill retention as once thought.

Finally, one of the objectives of this research was to use a more narrow, fine-grained classification of task types and characteristics than those that have been used in previous meta-analyses in hopes of providing answers to the question of what tasks are more or less susceptible to decay. Real-world tasks were split into categories based on the

occupations that the tasks were most relevant to (medical/dental, military, sports, transportation, miscellaneous) and tasks were classified based on their physical and psychomotor abilities (dexterity, gross-body coordination, precision/steadiness). Unfortunately, the findings were not particularly illuminating. The only category that had a notable influence on skill decay patterns was laboratory/artificial and moderator analyses for task ability demands were limited in their utility because there were high/low base rates associated with the ability demands of tasks that were measured within each performance measure category. The task ability demands findings suggest that perhaps tasks that involve gross body coordination and precision/steadiness are better retained over time than tasks that do not. But, given the low base rates for gross body coordination tasks in the retention literature, the lack of consistent findings across performance measure categories, and the lack of theoretical explanation for these findings, the results should be further investigated with follow-up studies before they are given substantial weight. Therefore, the question of which types of tasks are most susceptible to decay, other than laboratory/artificial tasks and less complex tasks, remains relatively unresolved.

6.4 Research Questions 10: Comparison of Different Effect Size Formulas

Based on the results of the Monte Carlo simulations, I decided to separate the effect sizes calculated using the independent samples formula from the other effect sizes and compare the effect sizes and skill retention trends to one another. The overall skill retention trends did not differ substantially between the independent samples formula effect sizes and the paired-samples formula/converted *t*-value effect sizes, which suggests that the decision to remove the effect sizes not calculated using the independent samples effect size formula likely did not influence the estimation of skill decline per month.

However, it was somewhat surprising that the overall magnitude of effect sizes not computed using the independent samples formula was lower after accounting for retention interval length in comparison to the overall magnitude of effect sizes computed using the independent samples formula because the Monte Carlo simulation findings suggested that they would likely be stronger. The smaller number of effect sizes not computed using the independent samples formula may partially account for this finding. But, it might also be that the paired t -values and Cohen's D statistics reported by authors represented multiple retention trials or multiple retention blocks, whereas I coded the means and SDs used in the independent samples formulas as only the first retention trial performance whenever possible. It is possible that the paired-samples formula/converted t -values were therefore more confounded by relearning effects than the independent samples formula effect sizes were, which may have led to an attenuation of the effect sizes reported by authors.

Another plausible explanation is that the paired t -values and Cohen's D statistics reported did not account for the correlation between acquisition and retention. As the Monte Carlo results suggested, the largest differences would be expected at high levels of correlation between measures. The effect sizes converted from paired samples t -tests (61 out of the 71 effect sizes not computed using the independent samples formula) did not account for the correlation between measures because it was rarely reported. It's possible that if the correlation between acquisition and retention had been more frequently reported and therefore used in the calculation of paired-sample effect sizes, the differences would have been more aligned with the simulation results.

6.5 Limitations and Future Research Directions

As is the case with any meta-analysis, the findings and the conclusions that can be drawn are limited by the nature of the literature in question. Recommendations for future research on the topic of skill retention naturally go hand in hand with the limitations of the literature. Both are discussed in this final section of the Discussion chapter in hopes that the discussion will stimulate work that addresses the primary gaps in the literature. One of the primary justifications for the need for this meta-analysis of skill retention was that previous meta-analyses lacked retention intervals of lengths that are reflective of skill retention in the real world (e.g., months and years). While there were more retention intervals of longer lengths to include, the literature still consists primarily of short retention intervals. Over one quarter of the retention intervals included in the primary analyses (meta-regression equations for effect sizes calculated using the independent samples formula) were two days or shorter and nearly half of them were one week or shorter. As a result, estimations of the extent to which skills decline at longer retention intervals are less robust than they would be if there were a larger number of eligible studies with longer retention intervals in the literature. While conducting longitudinal studies over months or years and finding participants that can refrain from using relevant skills present challenges, a larger number of methodologically sound studies with long retention intervals that report statistics that can be used to calculate effect sizes remains perhaps the skill retention literature's greatest need.

This research was also limited by the data reported in primary studies. The skill retention literature is interdisciplinary and different disciplines have different norms and protocols for data reporting. As a result, many studies did not include basic statistics needed for computing effect sizes (e.g., means, standard deviations/standard errors/confidence

intervals). It was not uncommon to come across studies that fit all other eligibility requirements, except the data consisted only of the number or proportion of participants who maintained performance over some proficiency level (i.e., pass/fail rates). While such statistics may be of practical use, they do not provide sufficient information to quantify the decrement that occurs when skills are not frequently used, because they may collapse across individuals with very different skill levels (e.g., an individual who barely passed the assessment and an individual with near expert performance). It was also not uncommon to see no descriptive statistics reported or to see central tendency data reported without variability data.

Even when relevant statistics were provided, they were often reported in a manner that made estimation of effect sizes difficult. Authors frequently reported variation data only as error bars on graphs, which can be used to visualize standard deviation, standard error of the mean, or confidence intervals – each of which can be used to calculate effect sizes. But because the numerical data were often not reported, rulers were required to measure and compute an estimation, which likely lowers reliability, especially when either the error bars or the performance metrics were small. Many studies were excluded because a confident estimation could not be obtained from error bars. Further, many sources provided such error bars or denoted variation as “+/-” some numeric value but did not indicate what metric of variation was reported. Inferences were made when possible, but many sources were excluded because it was not possible to determine whether the metrics represented standard deviations, standard errors, or confidence intervals. Future researchers should note that regardless of whether statistics are used to directly answer

their research questions, it is advantageous to report central tendency and variation data that represent performance at acquisition and performance at retention in numeric form.

Another limitation of the current study involved the lack of independence of observations in the analyses, which often was also a result of the data reported in primary studies. The proper computation of effect sizes for paired samples has been debated and there is not a single agreed-upon formula, but contemporary work suggests that dependency should be accounted for in effect size calculations by using a formula that includes the correlation between measures and the paired t -value (which relies on the standard deviation between difference scores) (Dunlap et al., 1996). As the Monte Carlo simulations presented in this dissertation suggest, the effect sizes yielded from this formula compared to the formula that does not account for the dependency of measures in skill retention studies can be quite different, especially when the correlations between acquisition and retention performance are high (which one would expect in most cases) and mean differences between acquisition and retention are large. Unfortunately, most studies did not report correlations between acquisition and retention or paired t -tests, which limited the ability of primary analyses to account for the dependence of measures. The field would greatly benefit if authors were to consistently report correlations between acquisition and retention paired t -statistics for all skill retention studies and account for them in their own effect size calculations as well.

The final set of recommendations is related to moderator results and potential confounding variables previously discussed that lend themselves to methodological considerations that future researchers should account for. Retention effect sizes in the meta-analysis were likely confounded by complex interactions between speed and accuracy

components of performance, as well as relearning effects. Future skill retention researchers should be mindful that speed and accuracy are not positively correlated within individuals and therefore should not be combined into a single measure of performance or investigated as the same level of a single moderator in future meta-analyses.

Additionally, task instructions should be tailored to the specific task and participants should be informed of what the primary goals are and how performance will be measured. If goals and performance measurement are not specified or if both speed and accuracy are emphasized, researchers may be inviting speed-accuracy-tradeoff confounds. Information on the instructions that were given to participants should also be explicitly reported. Whenever possible, trial-by-trial data should be reported in order to isolate relearning effects from retention estimates. Additionally, infrequent opportunities for performance during retention intervals should be either accounted for in the study design or transparently reported. If skill retention researchers are interested in reporting an estimate of skill retention, using multiple retention tests or participants who may have had access to performance opportunities through their work and/or education may not be ideal approaches. More appropriate approaches might involve using separate samples when using multiple retention intervals or using samples unlikely to encounter exposure to tasks that would require the skill in question. At the very least, the field would benefit if skill retention researchers were to consider the likelihood that their participants may have encountered performance opportunities through their education/work, measure the frequency of their exposure when possible, and report this information.

A final limitation relates to the mostly null moderator findings of real-world task types and task ability demands. The question of for which types of tasks performance is

most susceptible to decay, other than laboratory/artificial and low-complexity tasks, remains unknown. The field would greatly benefit from the development of a validated taxonomy of procedural skills because the dimensions of the taxonomy may lead to the identifications of task-related moderators that suggest which tasks are most susceptible and resistant to decay over intervals of non-use.

CHAPTER 7. CONCLUSIONS

The purpose of this research was to build upon prior empirical findings related to skill retention and to capitalize on the recent growth of the skill retention literature to conduct a meta-analysis focused specifically on the retention of procedural skills with a significant motor component. Conclusions that can be drawn from the findings of the meta-analysis are offered below:

7.1 State of the Skill Retention Literature

1. Experts on the topic of skill retention once noted that it was a not a particularly popular research topic (Adams, 1964; Farr 1987). The volume of this meta-analysis (457 sources), over 400 more than the Arthur et al. (1998) meta-analysis despite the narrower scope of only skills with a significant motor component, indicates that the field has come a long way not only since those authors' statements but even just within the most recent 20 years. Increased attention to the topic from areas outside of applied and experimental psychology (e.g., medicine, sports psychology, neuroscience) allowed for the increase in volume of studies included in this meta-analysis.
2. While the growth in literature has been substantial, there is still a need for more studies that use longer retention intervals (i.e., months and years rather than days or weeks). The majority of the studies in this meta-analysis used shorter retention intervals than would be typical of intervals of non-use in many occupations.

7.2 Skill Decay Function

3. Procedural skills decline as a linear function of retention interval length. This conclusion could not be drawn from prior meta-analyses because retention interval was treated as a categorical variable consisting of asymmetric categories of retention interval length. The use of meta-regression allowed for the treatment of retention interval as a continuous predictor, which allowed for the identification of a linear skill decay function.
4. On average, procedural skills declined at a rate of 0.08 SD units per month for accuracy-based performance and 0.06 per month for speed-based performance and performance that requires a mixture of accuracy and speed. Individuals in the 90th and 50th percentile of skill proficiency after initial acquisition experienced skill declines with infrequent use that placed them in the 61st and 16th percentiles, respectively, after 9 to 12 months for accuracy-based skills, 12 to 15 months for speed-based skills, and 6 to 9 months for mixed skills.
5. Initial skill acquisition gains were typically lost between one year and two and a half years after initial acquisition, if the skills were not used consistently.

7.3 Moderators of Skill Decline

Task-related and study-design variables not only moderate differences in performance between skill acquisition and skill retention as prior research suggests, but also moderate the rate at which skills decline over time.

6. Task type and complexity may be meaningful moderators of the rate of skill decline. Performance is better retained over time on real-world tasks and high-

complexity tasks than on laboratory/artificial and low-complexity tasks, respectively.

7. Performance opportunities and task instructions may be meaningful moderators of the rate of skill decline. Performance was better retained over time when infrequent performance opportunities were available during the retention interval and when task instructions were aligned with how performance was measured, respectively.

7.4 Potential Confounds of Skill Retention Estimates

8. Skill retention estimates may be substantially confounded by relearning effects. Prior research suggests that after periods of non-use, initial performance is low but improves rapidly. An accurate measure of retention performance consists of only the first retention trial. The fact that retention performance is often reported aggregated across retention trials or blocks may mean that skill decay effect sizes were underestimated.
9. Skill retention estimates may be confounded by Speed-Accuracy-Tradeoff functions. Speed and accuracy are not conceptually comparable in the context of skill acquisition because speed and accuracy are negatively related at the individual level. If an individual's accuracy-based performance declines over time, it is difficult to determine whether the performance decrement can be attributed to skill decline vs. a change in the individual's speed-accuracy tradeoff decision. Estimates may have been further confounded because researchers often combine speed and accuracy performance into a single aggregate measure of performance.

10. Skill retention estimates may be confounded by the lack of reporting or accounting for the correlation between acquisition and retention. Previous research and the results of the Monte Carlo simulations presented in this dissertation suggest that effect size estimates for paired samples are stronger when they account for the correlation between measures. The infrequency with which the correlations between acquisition and retention were reported may therefore have led to an underestimate of skill decay effect sizes.

In sum, the research presented in this dissertation advances skill retention research by focusing exclusively on procedural skills with a significant non-verbal motor component and by developing more precise estimates of skill decline after months or years of infrequent use than were possible in previous skill retention meta-analyses. A more comprehensive set of moderators than has previously been considered was tested and several potentially meaningful moderators were identified for future skill retention researchers to consider. Yet, opportunities remain ripe for future research. The types of real-world tasks that are more or less susceptible to decay remain unknown and there are opportunities to improve estimates of retention by isolating retention from relearning and by more carefully attending to complex interactions between speed and accuracy. It is my hope that this research draws attention to these issues, and perhaps a subsequent meta-analysis of procedural skill retention will develop even more accurate skill retention estimates as a result.

APPENDIX A. SEARCH TERMS FOR EACH DATABASE

When possible, nested Boolean logic was used. However, several databases did not handle nested Boolean logic well and separate searches were required. Search terms in several databases were further broadened or narrowed based on the volume of relevant results that initial searches yielded. A broader set of Boolean search terms (e.g., skill AND retention as well as retention AND procedural, rather than skill retention AND procedural) was entered into PsycINFO because the database generated a higher proportion of relevant studies than other databases. NOT terms were added to initial search terms in several databases in order to minimize the number of results returned that would not have fit the study's eligibility criteria (e.g., child, dementia).

Scopus	<p>ALL (skill AND (retention OR decay OR maintenance OR deterioration OR perishability OR degradation) AND (procedural OR psychomotor OR perceptual) AND NOT (ethanol OR dementia OR alzheimer* OR child* OR alcohol OR smoking OR therap* OR depress* OR synapse OR "social skills" OR animal*)) AND (EXCLUDE (SRCTYPE , "k") OR EXCLUDE (SRCTYPE , "b")) AND (EXCLUDE (DOCTYPE , "re") OR EXCLUDE (DOCTYPE , "bk") OR EXCLUDE (DOCTYPE , "no") OR EXCLUDE (DOCTYPE , "ed") OR EXCLUDE (DOCTYPE , "sh") OR EXCLUDE (DOCTYPE , "le") OR EXCLUDE (DOCTYPE , "tb") OR EXCLUDE (DOCTYPE , "dp") OR EXCLUDE (DOCTYPE , "er"))</p>
Google Scholar & ProQuest	<p>("skill retention" OR "skill decay" OR "skill deterioration" OR "skill perishability" or "skill maintenance" OR "skill degradation") AND ("procedural" OR "psychomotor" OR "perceptual")</p>

PubMed	skill and (retention or decay or deterioration or perishability or degradation) AND (procedural or psychomotor or perceptual) NOT (schizophrenia or animal* OR child* OR alzheimer* OR dementia OR ethanol OR alcohol OR smoking OR therap* OR depress* OR synapse OR "social skills")	
PsycINFO	skill AND retention skill AND decay skill AND deterioration skill AND maintenance skill AND perishability skill AND degradation procedural AND retention procedural AND decay procedural AND deterioration procedural AND maintenance procedural AND degradation	perceptual AND retention perceptual AND decay perceptual AND deterioration perceptual AND maintenance perceptual AND degradation psychomotor AND retention psychomotor AND decay psychomotor AND deterioration psychomotor AND maintenance psychomotor AND degradation
Web of Science	skill retention AND procedural skill retention AND perceptual skill retention AND psychomotor skill decay AND procedural skill decay AND perceptual skill decay AND psychomotor skill deterioration AND procedural skill deterioration AND perceptual skill deterioration AND psychomotor	skill perishability AND procedural skill perishability AND perceptual skill perishability AND psychomotor skill maintenance AND procedural skill maintenance and perceptual skill maintenance AND psychomotor skill degradation AND procedural skill degradation AND perceptual skill degradation AND psychomotor
National Technical Reports Library	"procedural skill" and (retention or decay)	"procedural knowledge" and (retention or decay)

APPENDIX B. TASK LIST

Medical/Dental Tasks	Laboratory/Artificial Tasks
Cardiopulmonary resuscitation	Tracking task
Surgical procedures	Balancing task
Tasks representing surgical skills (e.g., peg transfer)	Lever task
Cather insertion/central line placement	Key/keypad pressing
Ultrasound	Finger tapping
Donning/removing PPE, other medical equipment	Barrier knockdown
Intubation	Rotary pursuit/pursuit rotor
Airway management	Reaching task
Basic life support skills	Flexion/extension task
Dental procedures	Isometric force task
Lumbar puncture	Microswitch task
Hand washing	Bimanual coordination
Anesthesia skills	Timing task involving moving apparatus
Neonatal resuscitation	Mirror training
Spinal manipulation/chiropractic procedure	Tracing without a mirror
Intravenous cannulation	Ladder climbing
Advanced life support skills	Serial aiming task
Placement of intrauterine contraception	Speed cup stacking
Otoscopy clinical skills	Three segment timing task
Sopharyngeal swab taking	Light-based hand-eye coordination task
First aid	Blindfolded drawing
Treating status epilepticus	Dynamometer task
Head and skill examination	Timing task using stylus
Cervical examination	Precision walking
Simulated baby delivery with shoulder dystocia	Thumb abduction/acceleration
Electrocardiography	Index finger abductions
Bellocq's tamponade	Bilateral transfer task
Targeted temperature management	Movement distance recall task
Compounding capsule for prescription	Grooved pegboard task
Colonoscopy	Moving a mouse-operated cursor through arrays
Patient log roll and recovery position	Typing task
Emergency triage assessment and treatment	
Military Tasks	Sports Tasks
Flying a military aircraft	Golf

Gunner operation
 Using a gunner's quadrant
 Marksmanship/shooting a firearm
 Assembling/loading a firearm
 Launching/guiding a missile
 Individual field artillery tasks
 Advanced field artillery tasks
 M551 tank target engagement
 Start/stop tank engine
 Precision fire
 Ground guiding
 Target data display set up, use, and maintenance
 Use of tank inter-vehicular information system
 Diagnosing and troubleshooting turret malfunction
 Use of command post system
 Use of squad radio
 Underwater helicopter escape
 Simulated fire escape
 Egress skills

Tennis
 Darts
 Toss/throw ball
 Basketball Badminton
 Dance
 Volleyball
 Baseball/softball
 Bowling/cricket bowling
 Judo/martial arts
 Gymnastics
 Soccer/European football
 Australian rules football
 Hockey/field hockey
 Swimming
 Table tennis/ping pong
 Skiing/snowboarding
 Juggling
 Fencing
 Gaelic football
 Archery
 Cycling
 Rowing
 Weight lifting
 Lasso

Transportation Tasks

Flying a non-military aircraft
 Driving an automobile
 Manual destination entry task
 Simulated hydraulic excavator

Miscellaneous Tasks

Wheelchair operation
 Using a prosthetic
 Assembly task
 Video/computer games
 Menu-based computer task
 Basic computer skills
 Engine maintenance task
 Measurement task
 Tactile vernier task

APPENDIX C. LAB/ARTIFICIAL VS. REAL-WORLD TASK

MODERATOR ANALYSIS MODELS

Accuracy (k = 829)				
Model	β	<i>SE</i>	<i>t</i>	95% CI
Intercept	-0.13	0.05	-2.48*	[-0.22, -0.03]
Retention Interval	-0.16	0.02	-7.56**	[-0.21, -0.12]
Real-World Task	-0.20	0.06	-3.13**	[-0.32, -0.07]
Interaction	0.11	0.02	4.52**	[0.06, 0.15]
Tests				
Residual Heterogeneity	Moderators			
$Q(825) = 4070.26^{**}$	$F(3, 825) = 34.03^{**}$			
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.40	0.63	82.49%	5.71	17.91%

Note. * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes. The test of residual heterogeneity is Cochran's Q test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size. Laboratory/artificial was used as the reference group.

Speed (k = 267)				
Model	β	<i>SE</i>	<i>t</i>	95% CI
Intercept	0.06	0.09	0.70	[-0.11, 0.23]
Retention Interval	-0.11	0.05	-2.16*	[-0.21, -0.01]
Real-World Task	-0.32	0.11	-2.90**	[-0.53, -0.10]
Interaction	0.07	0.06	1.30	[-0.04, 0.19]
Tests				
Residual Heterogeneity	Moderators			
$Q(263) = 882.49^{**}$	$F(3, 263) = 5.67^{**}$			
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.34	0.58	71.79%	3.54	7.82%

Note. * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes. The test of residual heterogeneity is Cochran's Q test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size. Laboratory/artificial was used as the reference group.

APPENDIX D. ALL TASK TYPES MODERATOR ANALYSIS

MODELS

Accuracy (k = 829)				
Model	β	SE	<i>t</i>	95% CI
Intercept	-0.12	0.05	-2.50*	[-0.22, -0.03]
Retention Interval	-0.16	0.02	-7.70**	[-0.21, -0.12]
Task Type				
Medical/Dental	-0.41	0.08	-5.06**	[-0.56, -0.25]
M/T	-0.10	0.13	-0.70	[-0.35, 0.17]
Miscellaneous	0.07	0.16	0.41	[-0.25, 0.38]
Sports	-0.12	0.07	-1.57	[-0.26, 0.03]
Interaction Terms				
RI*Med/Den	0.12	0.02	4.88**	[0.07, 0.17]
RI*M/T	0.16	0.05	3.41**	[0.07, 0.24]
RI*Misc	0.13	0.04	3.47**	[0.06, 0.21]
RI*Sports	0.18	0.05	3.80**	[0.09, 0.27]
Tests				
Residual Heterogeneity	Moderators			
$Q(825) = 4015.93^{**}$	$F(9, 819) = 15.26^{**}$			
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.38	0.62	81.62%	5.44	21.49%

Note. * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes. The test of residual heterogeneity is Cochran's Q test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size. Laboratory/artificial was used as the reference group. M/T = military and transportation, Med/Den = medical/dental, Misc = miscellaneous.

Speed (k = 267)				
Model	β	<i>SE</i>	<i>t</i>	95% CI
Intercept	0.06	0.08	0.70	[-0.11, 0.23]
Retention Interval	-0.11	0.05	-2.18*	[-0.21, -0.01]
Task Type				
Medical/Dental	-0.40	0.12	-3.25**	[-0.64, -0.16]
M/R	-0.07	0.26	-0.26	[-0.57, 0.43]
Miscellaneous	-0.37	0.16	-2.31*	[-0.68, -0.06]
Sports	0.67	1.33	0.50	[-1.95, 3.28]
Interaction Terms				
RI*Med/Den	0.08	0.06	1.42	[-0.03, 0.20]
RI*M/T	-0.03	0.11	-0.21	[-0.24, 0.17]
RI*Misc	0.42	0.14	2.96**	[0.14, 0.70]
RI*Sports	-2.05	3.01	-0.68	[-7.97, 3.87]
Tests				
Residual Heterogeneity	Moderators			
$Q(257) = 851.94^{**}$	$F(9, 257) = 3.19^{**}$			
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.33	0.58	70.91%	3.44	10.37%

Note. * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes. The test of residual heterogeneity is Cochran's Q test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size. Laboratory/artificial was used as the reference group. M/T = military and transportation, Med/Den = medical/dental, Misc = miscellaneous.

APPENDIX E. OPEN VS. CLOSED-LOOPED TASK MODERATOR

ANALYSIS MODELS

Accuracy (k = 829)				
Model	β	<i>SE</i>	<i>t</i>	95% CI
Intercept	-0.29	0.04	-7.04**	[-0.37, -0.21]
Retention Interval	-0.08	0.01	-7.08**	[-0.10, -0.06]
Open-Looped	0.09	0.06	1.50	[-0.03, 0.22]
Interaction	0.01	0.02	0.45	[-0.03, 0.04]
Tests				
Residual Heterogeneity	Moderators			
$Q(825) = 4230.11^{**}$	$F(3, 825) = 26.80^{**}$			
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.42	0.65	82.03%	5.89	14.41%

Note. * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes. The test of residual heterogeneity is Cochran's Q test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size. Closed-looped was used as the reference group.

Speed (k = 267)				
Model	β	<i>SE</i>	<i>t</i>	95% CI
Intercept	-0.22	0.08	-2.84**	[-0.38, -0.07]
Retention Interval	-0.09	0.03	-2.80**	[-0.16, -0.03]
Open-Looped	0.18	0.11	1.70	[-0.03, 0.39]
Interaction	0.06	0.04	1.30	[-0.03, 0.14]
Tests				
Residual Heterogeneity	Moderators			
$Q(263) = 885.78^{**}$	$F(3, 263) = 6.07^{**}$			
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.34	0.58	71.65%	3.53	8.40%

Note. * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes. The test of residual heterogeneity is Cochran's Q test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size. Closed-looped was used as the reference group.

Mixed (k = 180)				
Model	β	<i>SE</i>	<i>t</i>	95% CI
Intercept	-0.14	0.12	-1.24	[-0.38, 0.09]
Retention Interval	-0.14	0.02	-5.75**	[-0.19, -0.09]
Open-Looped	-0.66	0.15	-4.53**	[-0.95, -0.38]
Interaction	0.11	0.03	3.80**	[0.05, 0.16]
Tests				
Residual Heterogeneity		Moderators		
$Q(176) = 877.95^{**}$		$F(3, 176) = 16.04^{**}$		
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.34	0.58	71.65%	3.53	8.40%

Note. * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes. The test of residual heterogeneity is Cochran's Q test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size. Closed-looped was used as the reference group.

APPENDIX F. COMPONENT COMPLEXITY TASK MODERATOR

ANLAYSIS MODELS

Accuracy (k = 829)				
Model	β	<i>SE</i>	<i>t</i>	95% CI
Intercept	-0.11	0.08	-1.44	[-0.27, 0.04]
Retention Interval	-0.13	0.02	-6.37**	[-0.17, -0.09]
Component	-0.08	0.04	-1.87	[-0.17, 0.00]
Interaction	0.03	0.01	2.92**	[0.01, 0.05]
Tests				
Residual Heterogeneity	Moderators			
$Q(825) = 4193.17^{**}$	$F(3, 825) = 28.66^{**}$			
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.41	0.64	82.85%	5.83	15.31%

Note. * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes. The test of residual heterogeneity is Cochran's Q test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size.

Speed (k = 267)				
Model	β	<i>SE</i>	<i>t</i>	95% CI
Intercept	0.14	0.12	1.19	[-0.09, 0.37]
Retention Interval	-0.10	0.06	-1.72	[-0.22, 0.01]
Component	-0.16	0.06	-2.68**	[-0.28, -0.04]
Interaction	0.02	0.02	0.91	[-0.03, 0.07]
Tests				
Residual Heterogeneity	Moderators			
$Q(263) = 879.24^{**}$	$F(3, 263) = 5.31^{**}$			
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.34	0.58	71.82%	3.55	7.63%

Note. * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes. The test of residual heterogeneity is Cochran's Q test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size.

Mixed (k = 180)				
Model	β	<i>SE</i>	<i>t</i>	95% CI
Intercept	0.31	0.19	1.66	[-0.06, 0.68]
Retention Interval	-0.21	0.04	-5.66**	[-0.29, -0.14]
Component	-0.40	0.08	-5.04**	[-0.55, -0.24]
Interaction	0.06	0.02	4.28**	[0.03, 0.09]
Tests				
Residual Heterogeneity	Moderators			
$Q(176) = 844.75^{**}$	$F(3, 176) = 18.16^{**}$			
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.35	0.59	81.55%	5.42	32.95%

Note. * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes. The test of residual heterogeneity is Cochran's Q test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size.

APPENDIX G. COORDINATIVE COMPLEXITY TASK

MODERATOR ANALYSIS MODELS

Accuracy (k = 829)				
Model	β	<i>SE</i>	<i>t</i>	95% CI
Intercept	-0.23	0.09	-2.60**	[-0.40, -0.06]
Retention Interval	-0.12	0.03	-4.05**	[-0.17, -0.06]
Coordinative	-0.01	0.04	-0.22	[-0.09, 0.07]
Interaction	0.02	0.01	1.43	[-0.01, 0.04]
Tests				
Residual Heterogeneity	Moderators			
$Q(825) = 4230.65^{**}$	$F(3, 825) = 26.20^{**}$			
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.42	0.65	83.07%	5.91	13.97%

Note. * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes. The test of residual heterogeneity is Cochran's Q test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size.

Speed (k = 267)				
Model	β	<i>SE</i>	<i>t</i>	95% CI
Intercept	-0.02	0.16	-0.11	[-0.32, 0.29]
Retention Interval	0.08	0.07	-1.18	[-0.22, 0.06]
Coordinative	0.06	0.07	-0.81	[-0.19, 0.08]
Interaction	0.01	0.03	0.32	[-0.05, 0.06]
Tests				
Residual Heterogeneity	Moderators			
$Q(263) = 893.93^{**}$	$F(3, 263) = 2.99^*$			
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.35	0.60	72.51%	3.64	4.25%

Note. * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes. The test of residual heterogeneity is Cochran's Q test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size.

Mixed (k = 180)				
Model	β	<i>SE</i>	<i>t</i>	95% CI
Intercept	0.38	0.23	1.68	[-0.07, 0.83]
Retention Interval	-0.24	0.06	-4.05**	[-0.36, -0.12]
Coordinative	-0.38	0.09	-4.41**	[-0.56, -0.21]
Interaction	0.07	0.02	3.17**	[0.03, 0.11]
Tests				
Residual Heterogeneity	Moderators			
$Q (176) = 901.00^{**}$	$F (3, 176) = 15.43^{**}$			
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.37	0.61	83.22%	5.96	27.76%

Note. * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes. The test of residual heterogeneity is Cochran's Q test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size.

APPENDIX H. DYNAMIC COMPLEXITY TASK MODERATOR

ANALYSIS MODELS

Accuracy (k = 829)				
Model	β	<i>SE</i>	<i>t</i>	95% CI
Intercept	-0.16	0.07	-2.23*	[-0.30, -0.02]
Retention Interval	-0.12	0.02	-5.30**	[-0.16, -0.07]
Dynamic	-0.06	0.04	-1.39	[-0.14, 0.02]
Interaction	0.02	0.01	1.91	[0.00, 0.05]
Tests				
Residual Heterogeneity	Moderators			
$Q(825) = 4216.33^{**}$	$F(3, 825) = 26.86^{**}$			
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.42	0.65	83.03%	5.89	14.39%

Note. * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes. The test of residual heterogeneity is Cochran's Q test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size.

Speed (k = 267)				
Model	β	<i>SE</i>	<i>t</i>	95% CI
Intercept	0.11	0.12	0.89	[-0.13, 0.35]
Retention Interval	-0.15	0.06	-2.30*	[-0.27, -0.02]
Dynamic	-0.14	0.06	-2.21*	[-0.26, -0.02]
Interaction	0.04	0.03	1.53	[-0.01, 0.09]
Tests				
Residual Heterogeneity	Moderators			
$Q(263) = 892.44^{**}$	$F(3, 263) = 4.57^*$			
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.35	0.59	72.02%	3.57	6.64%

Note. * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes. The test of residual heterogeneity is Cochran's Q test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size.

Mixed (k = 180)				
Model	β	<i>SE</i>	<i>t</i>	95% CI
Intercept	0.35	0.19	1.80	[-0.03, 0.73]
Retention Interval	-0.21	0.04	-4.82**	[-0.29, -0.12]
Dynamic	-0.41	0.08	-5.09**	[-0.57, -0.25]
Interaction	0.06	0.02	3.64**	[0.03, 0.09]
Tests				
Residual Heterogeneity	Moderators			
$Q(176) = 858.77^{**}$	$F(3, 176) = 17.85^{**}$			
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.35	0.60	82.66%	5.77	31.66%

Note. * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes. The test of residual heterogeneity is Cochran's Q test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size.

APPENDIX I. DEXTERITY MODERATOR ANALYSIS MODELS

Accuracy (k = 829)				
Model	β	SE	<i>t</i>	95% CI
Intercept	-0.37	0.09	-4.05**	[-0.55, -0.19]
Retention Interval	-0.18	0.09	-1.96	[-0.35, 0.00]
Dexterity (Y)	0.15	0.10	1.54	[-0.04, 0.34]
Interaction	0.10	0.09	1.07	[-0.08, 0.27]
Tests				
Residual Heterogeneity	Moderators			
$Q(825) = 4232.89^{**}$	$F(3, 825) = 28.07^{**}$			
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.42	0.65	83.11%	5.92	13.78%

Note. * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes. The test of residual heterogeneity is Cochran's Q test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size. "No" dexterity was chosen as the reference group.

Speed (k = 267)				
Model	β	SE	<i>t</i>	95% CI
Intercept	0.07	0.16	0.44	[-0.24, 0.38]
Retention Interval	0.02	0.09	0.21	[-0.16, 0.20]
Dexterity (Y)	-0.24	0.17	-1.43	[-0.57, 0.09]
Interaction	-0.08	0.10	-0.89	[-0.27, 0.10]
Tests				
Residual Heterogeneity	Moderators			
$Q(263) = 891.52^{**}$	$F(3, 263) = 4.87^*$			
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.35	0.59	71.96%	3.57	6.57%

Note. * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes. The test of residual heterogeneity is Cochran's Q test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size. "No" dexterity was chosen as the reference group.

Mixed (k = 180)				
Model	β	SE	<i>t</i>	95% CI
Intercept	-1.88	0.74	-3.34*	[-3.34, -0.42]

Retention Interval	0.19	0.12	-0.05	[-0.05, 0.42]
Dexterity (Y)	1.31	0.74	-0.16	[-0.16, 2.78]
Interaction	-0.25	0.12	-0.49*	[-0.49, -0.02]
Tests				
Residual Heterogeneity	Moderators			
$Q(176) = 1023.14^{**}$	$F(3, 176) = 9.41^{**}$			
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.43	0.66	85.55%	6.92	17.22%

Note. * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes. The test of residual heterogeneity is Cochran's Q test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size. "No" dexterity was chosen as the reference group.

APPENDIX J. GROSS BODY COORDINATION MODERATOR

ANLAYSIS MODELS

Accuracy (k = 829)				
Model	β	<i>SE</i>	<i>t</i>	95% CI
Intercept	-0.28	0.04	-7.31**	[-0.35, -0.20]
Retention Interval	-0.08	0.01	-8.30**	[-0.10, -0.06]
Gross Body (Y)	0.07	0.07	1.03	[-0.06, 0.20]
Interaction	0.09	0.04	0.04*	[0.01, 0.18]
Tests				
Residual Heterogeneity	Moderators			
$Q(825) = 4233.76^{**}$	$F(3, 825) = 28.11^{**}$			
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.42	0.64	83.05%	5.90	14.51%

Note. * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes. The test of residual heterogeneity is Cochran's *Q* test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size. "No" gross body coordination was used as the reference group.

Speed (k = 267)				
Model	β	<i>SE</i>	<i>t</i>	95% CI
Intercept	-0.22	0.06	-3.90**	[-0.33, -0.11]
Retention Interval	-0.06	0.02	-2.62**	[-0.20, -0.01]
Gross Body (Y)	0.56	0.15	3.62**	[0.25, 0.86]
Interaction	0.02	0.09	0.24	[-0.16, 0.20]
Tests				
Residual Heterogeneity	Moderators			
$Q(263) = 861.72^{**}$	$F(3, 263) = 9.30^*$			
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.32	0.57	70.67%	3.41	12.29%

Note. * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes. The test of residual heterogeneity is Cochran's *Q* test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size. "No" gross body coordination was used as the reference group.

Mixed (k = 180)				
Model	β	<i>SE</i>	<i>t</i>	95% CI
Intercept	-0.59	0.08	-7.86**	[-0.74, -0.44]
Retention Interval	-0.06	0.01	-4.81**	[-0.09, -0.04]
Gross Body (Y)	0.29	0.50	0.58	[-0.70, 1.29]
Interaction	0.02	0.09	0.17	[-0.16, 0.19]
Tests				
Residual Heterogeneity	Moderators			
$Q (176) = 1039.76^{**}$	$F (3, 176) = 8.07^{**}$			
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.44	0.66	85.90%	7.09	14.81%

Note. * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes. The test of residual heterogeneity is Cochran's Q test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size. "No" gross body coordination was used as the reference group.

APPENDIX K. PRECISION/STEADINESS MODERATOR

ANLAYSIS MODELS

Accuracy (k = 829)				
Model	β	<i>SE</i>	<i>t</i>	95% CI
Intercept	-0.19	0.06	-3.33**	[-0.31, -0.08]
Retention Interval	-0.08	0.01	-5.81**	[-0.11, -0.06]
Precision (Y)	-0.08	0.07	-1.18	[-0.22, 0.05]
Interaction	0.01	0.02	0.35	[-0.09, 0.04]
Tests				
Residual Heterogeneity	Moderators			
$Q(825) = 4232.18^{**}$	$F(3, 825) = 25.92^{**}$			
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.42	0.65	83.08%	5.91	13.70%

Note. * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes. The test of residual heterogeneity is Cochran's Q test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size. "No" precision/steadiness was used as the reference group.

Speed (k = 267)				
Model	β	<i>SE</i>	<i>t</i>	95% CI
Intercept	0.02	0.08	0.19	[-0.15, 0.18]
Retention Interval	-0.07	0.04	-1.79	[-0.14, 0.01]
Precision (Y)	-0.26	0.11	-2.43*	[-0.48, -0.05]
Interaction	0.01	0.05	0.20	[-0.08, 0.10]
Tests				
Residual Heterogeneity	Moderators			
$Q(263) = 861.72^{**}$	$F(3, 263) = 9.30^*$			
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.32	0.57	70.67%	3.41	12.29%

Note. * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes. The test of residual heterogeneity is Cochran's Q test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size. "No" precision/steadiness was used as the reference group.

Mixed (k = 180)				
Model	β	<i>SE</i>	<i>t</i>	95% CI
Intercept	-0.14	0.17	-0.82	[-0.47, 0.19]
Retention Interval	-0.14	0.03	-4.89**	[-0.20, -0.08]
Precision (Y)	-0.53	0.19	-2.86**	[-0.90, -0.16]
Interaction	0.10	0.03	3.01**	[0.03, 0.16]
Tests				
Residual Heterogeneity	Moderators			
$Q(176) = 902.25^{**}$	$F(3, 176) = 11.52^{**}$			
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.40	0.63	83.22%	5.96	22.96%

Note. * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes. The test of residual heterogeneity is Cochran's Q test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size. "No" precision/steadiness was used as the reference group.

APPENDIX L. PERFORMANCE OPPORTUNITY MODERATOR

ANALYSIS MODELS

Accuracy (k = 829)				
Model	β	<i>SE</i>	<i>t</i>	95% CI
Intercept	-0.21	0.03	-6.53**	[-0.27, -0.14]
Retention Interval	-0.08	0.01	-8.28**	[-0.10, -0.06]
Perf. Opp. (Y)	-0.59	0.12	-5.00**	[-0.82, -0.36]
Interaction	0.06	0.02	2.91**	[0.02, 0.11]
Tests				
Residual Heterogeneity	Moderators			
$Q(825) = 4155.40^{**}$	$F(3, 825) = 34.53^{**}$			
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.41	0.64	82.69%	5.78	15.24%

Note. * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes, Perf. Opp. = performance opportunity. The test of residual heterogeneity is Cochran's Q test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size. "No" performance opportunity was used as the reference group.

Speed (k = 267)				
Model	β	<i>SE</i>	<i>t</i>	95% CI
Intercept	-0.14	0.06	-2.56*	[-0.25, -0.03]
Retention Interval	-0.08	0.03	-3.00**	[-0.13, -0.03]
Perf. Opp. (Y)	0.19	0.20	0.91	[-0.22, 0.59]
Interaction	0.02	0.05	0.42	[-0.08, 0.13]
Tests				
Residual Heterogeneity	Moderators			
$Q(263) = 901.37^{**}$	$F(3, 263) = 3.66^{**}$			
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.35	0.59	72.48%	3.43	4.79%

Note. * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes, Perf. Opp. = performance opportunity. The test of residual heterogeneity is Cochran's Q test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size. "No" performance opportunity was used as the reference group.

Mixed (k = 180)				
Model	β	<i>SE</i>	<i>t</i>	95% CI
Intercept	-0.55	0.08	-7.21***	[-0.71, -0.40]
Retention Interval	-0.08	0.02	-5.24**	[-0.12, -0.05]
Perf. Opp. (Y)	-0.19	0.27	-0.70	[-0.71, 0.34]
Interaction	0.06	0.03	1.88	[0.00, 0.13]
Tests				
Residual Heterogeneity	Moderators $F(3, 176) = 19.62^{**}$			
	$Q(176) = 947.91^{**}$			
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.42	0.65	84.42%	6.42	18.24%

Note. * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes, Perf. Opp. = performance opportunity. The test of residual heterogeneity is Cochran's Q test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size. "No" performance opportunity was used as the reference group.

APPENDIX M. TASK INSTRUCTIONS MODERATOR ANALYSIS

MODELS

Accuracy (k = 829)				
Model	β	SE	<i>t</i>	95% CI
Intercept	-0.20	0.04	-5.48**	[-0.28, -0.13]
Retention Interval	-0.08	0.01	-6.05**	[-0.11, -0.05]
Instructions				
Speed/Mixed	0.19	0.10	1.96*	[0.00, 0.38]
Not Clear	-0.31	0.08	-4.07**	[-0.46, -0.16]
Interaction				
RI*Speed/Mix	-0.08	0.04	-2.12*	[-0.16, -0.01]
RI*Not Clear	0.04	0.02	2.09*	[0.00, 0.08]
Tests				
Residual Heterogeneity	Moderators			
$Q(825) = 4136.87^{**}$	$F(3, 825) = 20.93^{**}$			
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.40	0.63	82.48%	5.71	17.46%

Note. * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes, RI = retention interval, Speed/Mixed = Speed or mixed. The test of residual heterogeneity is Cochran's Q test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size. Accuracy instructions was chosen as the reference group.

Speed (k = 267)				
Model	β	<i>SE</i>	<i>t</i>	95% CI
Intercept	-0.10	0.14	-0.70	[-0.37, 0.17]
Retention Interval	0.01	0.07	0.09	[-0.14, 0.15]
Instructions				
Speed or Mixed	0.05	0.16	0.34	[-0.26, 0.37]
Not Clear	-0.17	0.16	-1.04	[-0.49, 0.15]
Interaction				
RI*Acc or Mix	-0.11	0.08	-1.37	[-0.26, 0.05]
RI*Not Clear	-0.04	0.08	-0.45	[-0.19, 0.12]
Tests				
Residual Heterogeneity		Moderators		
$Q (825) = 880.94^{**}$		$F (3, 825) = 2.83^*$		
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.35	0.60	72.07%	3.58	5.90%

Note. * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes, RI = retention interval. The test of residual heterogeneity is Cochran's Q test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size. Accuracy instructions was chosen as the reference group.

Mixed (k = 180)				
Model	β	<i>SE</i>	<i>t</i>	95% CI
Intercept	-0.39	0.13	-3.01	[-0.65, -0.14]
Retention Interval	-0.04	0.05	-0.81	[-0.13, 0.05]
Instructions				
Just Acc or Speed	0.53	0.36	1.48	[-0.18, 1.25]
Not Clear	-0.37	0.16	-2.36	[-0.69, -0.06]
Interaction				
RI* Acc/Spe	-0.22	0.14	-1.61	[-0.50, 0.05]
RI*Not Clear	-0.01	0.05	-0.25	[-0.10, 0.08]
Tests				
Residual Heterogeneity	Moderators			
$Q(825) = 1000.36^{**}$	$F(3, 825) = 7.69^{**}$			
Heterogeneity Statistics				
τ^2	τ	I^2	H^2	R^2
0.41	0.64	84.57%	6.44	20.47%

Note. * < 0.05, ** < 0.01. *Note.* * < 0.05, ** < 0.01. SE = standard error, CI = confidence interval, k = number of effect sizes, RI = retention interval, Acc= accuracy, Acc/spe = Just accuracy or speed. The test of residual heterogeneity is Cochran's Q test, which is a test of whether there is significant heterogeneity remaining after model predictors have been accounted for. The test of moderators is a Wald test of whether the moderators in the model (i.e., all predictors) significantly predict the effect size. Accuracy instructions was chosen as the reference group.

REFERENCES

- Ackerman, P. L. (1992). Predicting individual differences in complex skill acquisition: Dynamics of ability determinants. *Journal of Applied Psychology, 77*(5), 598–614. <https://doi.org/10.1037/0021-9010.77.5.598>
- Ackerman, P. L. (2008). Knowledge and cognitive aging. In Craik, F.I. & Salthouse, T.A. (Eds) *The handbook of aging and cognition* (pp.445-489). New York: Psychology Press.
- *Ackermann, A. (2007). *Acquisition and retention of CPR knowledge and skills for junior level baccalaureate nursing students*. [Doctoral dissertation, Duquesne University]. Duquesne Scholarship Collection. <https://dsc.duq.edu/etd/13>
- Adams, J. A. (1964). Motor skills. *Annual Review of Psychology, 15*(1), 181-202. <https://doi.org/10.1146/annurev.ps.15.020164.001145>
- Adams, J. A. (1967). *Human memory*. New York: McGraw-Hill
- Adams, J. A. (1971). A closed-loop theory of motor learning. *Journal of Motor Behavior, 3*(2), 111-150. <https://doi.org/10.1080/00222895.1971.10734898>
- Adams, J. A. (1987). Historical review and appraisal of research on the learning, retention, and transfer of human motor skills. *Psychological Bulletin, 101*(1), 41–74. <https://doi.org/10.1037/0033-2909.101.1.41>
- *Ahn, J., Yashar, M. D., Novack, J., Davidson, J., Lapin, B., Ocampo, J., & Wang, E. (2016). Mastery learning of video laryngoscopy Using the glidescope in the emergency department. *Simulation in Healthcare, 11*(5), 309-315. <https://doi.org/10.1097/sih.0000000000000164>
- *de Albuquerque, L. L., Fischer, K. M., Pauls, A. L., Pantovic, M., Guadagnoli, M. A., Riley, Z. A., & Poston, B. (2019). An acute application of transcranial random noise stimulation does not enhance motor skill acquisition or retention in a golf putting task. *Human Movement Science, 66*, 241-248. <https://doi.org/10.1016/j.humov.2019.04.017>

- *Aljamal, Y., Cook, D. A., Sedlack, R. E., Kelley, S. R., & Farley, D. R. (2019). An Inexpensive, Portable Physical Endoscopic Simulator: Description and Initial Evaluation. *Journal of Surgical Research*, 243, 560-566. <https://doi.org/10.1016/j.jss.2019.07.012>
- *Allnutt, M. F., & Everhart, C. D. (1980). *Retraining Army aviators following a protracted absence from flying* (Research Report 1296). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. <https://apps.dtic.mil/sti/citations/tr/ADA117548>
- *Almeida, D., Clark, C., Jones, M., McConnell, P., & Williams, J. (2020). Consistency and variability in human performance during simulate infant CPR: a reliability study. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, 28(1), 91. <https://doi.org/10.1186/s13049-020-00785-y>
- *Andrew, D., Yields, P., & Murphy, B. (2015). Do pursuit movement tasks lead to differential changes in early somatosensory evoked potentials related to motor learning compared with typing tasks? *Journal of Neurophysiology*, 113(4), 1156-1164. <https://doi.org/10.1152/jn.00713.2014>
- *Aqel, A. A., & Ahmad, M. M. (2014). High-fidelity simulation effects on CPR knowledge, skills, acquisition, and retention in nursing students. *Worldviews on Evidence-Based Nursing*, 11(6), 394-400. <https://doi.org/10.1111/wvn.12063>
- *Arain, M., Begum, S., Shariff, A., Khan, S., Pal, K. I., Khan, M. R., Ali, M., & Ringers, J. (2022). Prospective comparison of single encounter versus distributed laparoscopic training in novice learners: A controlled trial. *Journal of Education and Health Promotion*, 11(1). https://doi.org/10.4103/jehp.jehp_825_21
- *Aranda-García, S., Herrera-Pedroviejo, E., & Abelairas-Gómez, C. (2019). Basic life-support learning in undergraduate students of sports sciences: Efficacy of 150 minutes of training and retention after eight months. *International Journal of Environmental Research and Public Health*, 16(23), 4771. <https://doi.org/10.3390/ijerph16234771>
- *Arlington, L., Kairuki, A. K., Isangula, K. G., Meda, R. A., Thomas, E., Temu, A., Mponzi, V., Bishanga, D., Msemu, G., Azayo, M., & Nelson, B. D. (2017). Implementation of "Helping Babies Breathe": A 3-year experience in Tanzania. *Pediatrics*, 139(5). <https://doi.org/10.1542/peds.2016-2132>

- Arthur Jr., W., Bennett Jr, W., Stanush, P. L., & McNelly, T. L. (1998). Factors that influence skill decay and retention: A quantitative review and analysis. *Human Performance, 11*(1), 57-101. https://doi.org/10.1207/s15327043hup1101_3
- *Arnesen, K. A., Frithioff, A., Sørensen, M. S., Andersen, S. A. W., & Frendø, M. (2022). Mastoidectomy Training: Is Anatomical Variation Needed? A Randomized, Controlled Trial on Performance and Skills Transfer From Virtual Reality to a Three-Dimensional Printed Model. *Otology & Neurotology, 43*(8), 900-907. <https://doi.org/10.1097/mao.0000000000003607>
- *Arntfield, R., Pace, J., McLeod, S., Granton, J., Hegazy, A., & Lingard, L. (2015). Focused transesophageal echocardiography for emergency physicians-description and results from simulation training of a structured four-view examination. *Critical Ultrasound Journal, 7*, 10. <https://doi.org/10.1186/s13089-015-0027-3>
- *Atkins, R. J., Landsdowne, A. T. G., Pfister, H. P., & Provost, S. C. (2002). Conversion between control mechanisms in simulated flight: An ab initio quasi-transfer study. *Australian Journal of Psychology, 54*(3), 144-149. <https://doi.org/10.1080/00049530412331312714>
- *Awad-Igbaria, Y. (2019). *Motor learning in young adults: Is teacher attitude a factor in skill acquisition and motor memory consolidation?* (Publication No. 28745577). [Doctoral dissertation, University of Haifa]. ProQuest Dissertations & Theses Global.
- *Baker, C. H. (1968). An evaluation of guidance in learning a motor skill. *Canadian Journal of Psychology/Revue canadienne de psychologie, 22*(3), 217-227. <https://doi.org/10.1037/h0082762>
- *Badets, A., Blandin, Y., & Shea, C. H. (2006). Intention in motor learning through observation. *Quarterly Journal of Experimental Psychology, 59*(2), 377-386. <https://doi.org/10.1080/02724980443000773>
- *Balhaddad, A. A., Marghalani, A. A., Raderman, M. A., Miginsky, E., Massey, W., Strassler, H., & Anne Melo, M. (2021). Hands-on training based on quantifying radiant exposure improves how dental students cure composites: Skill retention at 2-year follow-up. *European Journal of Dental Education, 25*(3), 582-591. <https://doi.org/10.1111/eje.12635>
- *Barsuk, J. H., Cohen, E. R., McGaghie, W. C., & Wayne, D. B. (2010). Long-term retention of central venous catheter insertion skills after simulation-based mastery

learning. *Academic Medicine*, 85(10 Suppl), S9-12.
<https://doi.org/10.1097/ACM.0b013e3181ed436c>

- *Batista, M. T. S., Figueiredo, L. S., Martins, C. A., Nogueira, N. G. D. H. M., Ferreira, B. D. P., & Benda, R. N. (2022). Expectation of teaching and self-controlled KR in motor skills acquisition: Are there additive effects? *Research Quarterly for Exercise and Sport*. <https://doi.org/10.1080/02701367.2021.1999891>

- *Behzadnia, B., Mohammadzadeh, H., & Ahmadi, M. (2019). Autonomy-supportive behaviors promote autonomous motivation, knowledge structures, motor skills learning and performance in physical education. *Current Psychology: A Journal for Diverse Perspectives on Diverse Psychological Issues*, 38(6), 1692-1705.
<https://doi.org/10.1007/s12144-017-9727-0>

- *Berghuis, K. M. M., De Rond, V., Zijdwind, I., Koch, G., Veldman, M. P., & Hortobágyi, T. (2016). Neuronal mechanisms of motor learning are age dependent. *Neurobiology of Aging*, 46, 149-159.
<https://doi.org/10.1016/j.neurobiolaging.2016.06.013>

- *Berghuis, K. M. M., Fagioli, S., Maurits, N. M., Zijdwind, I., Marsman, J. B. C., Hortobágyi, T., Koch, G., & Bozzali, M. (2019). Age-related changes in brain deactivation but not in activation after motor learning. *NeuroImage*, 186, 358-368.
<https://doi.org/10.1016/j.neuroimage.2018.11.010>

- *Bested, S. R., de Grosbois, J., Crainic, V. A., & Tremblay, L. (2019). The influence of robotic guidance on error detection and correction mechanisms. *Human Movement Science*, 66, 124-132. <https://doi.org/10.1016/j.humov.2019.03.009>

- *Bickenbach, J., Schälte, G., Beckers, S., Fries, M., Derwall, M., & Rossaint, R. (2009). The intuitive use of laryngeal airway tools by first year medical students. *BMC Emergency Medicine*, 9, 18. <https://doi.org/10.1186/1471-227x-9-18>

- *Bieńkiewicz, M. M. N., Bringoux, L., Buloup, F., Rodger, M., Craig, C., & Bourdin, C. (2019). The limitations of being a copycat: Learning golf putting through auditory and visual guidance. *Frontiers in Psychology*, 10, 92.
<https://doi.org/10.3389/fpsyg.2019.00092>

- *Bink, M. L., Wampler, R. L., & Cage, E. A. (2011). *The retention of digital skills: Command post of the future* (Research Report 1937). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
<https://apps.dtic.mil/sti/citations/ADA535536>

- *Birrenbach, T., Zbinden, J., Papagiannaki, G., Exadaktylos, A. K., Müller, M., Hautz, W. E., & Sauter, T. C. (2021). Effectiveness and utility of virtual reality simulation as an educational tool for safe performance of covid-19 diagnostics: Prospective, randomized pilot trial. *JMIR Serious Games*, 9(4), e29586. <https://doi.org/10.2196/29586>
- *Bitsika, E., Karlis, G., Iacovidou, N., Georgiou, M., Kontodima, P., Vardaki, Z., & Xanthos, T. (2014). Comparative analysis of two venipuncture learning methods on nursing students. *Nurse Education Today*, 34(1), 15-18. <https://doi.org/10.1016/j.nedt.2013.03.016>
- *Bjerrum, A. S., Eika, B., Charles, P., & Hilberg, O. (2014). Dyad practice is efficient practice: A randomised bronchoscopy simulation study. *Medical Education*, 48(7), 705-712. <https://doi.org/10.1111/medu.12398>
- *Bjerrum, A. S., Eika, B., Charles, P., & Hilberg, O. (2016). Distributed practice. The more the merrier? A randomised bronchoscopy simulation study. *Medical Education Online*, 21(1), 30517. <https://doi.org/10.3402/meo.v21.30517>
- Bodilly, S., Fernandez, J., Kimbrough, J., & Purnell, S. (1986). *Individual Ready Reserve Skill Retention and Refresher Training Options* (RAND Note N-2535-RA). Santa Monica: RAND Corporation. <https://apps.dtic.mil/sti/citations/ADA183416>
- Boet, S., Borges, B. C. R., Naik, V. N., Siu, L. W., Riem, N., Chandra, D., ... & Joo, H. S. (2011). Complex procedural skills are retained for a minimum of 1 yr after a single high-fidelity simulation training session. *British Journal of Anaesthesia*, 107(4), 533-539. <https://doi.org/10.1093/bja/aer160>
- *Bomholt, K. B., Krogh, L. Q., Bomholt, S. R., Nebbjerg, M. A., Thim, T., & Løfgren, B. (2019). Three-month retention of basic life support with an automated external defibrillator using a two-stage versus four-stage teaching technique. *BioMed Research International*, 2019, 1394972. <https://doi.org/10.1155/2019/1394972>
- *Bonassi, G., Lagravinese, G., Bisio, A., Ruggeri, P., Pelosin, E., Bove, M., & Avanzino, L. (2020). Consolidation and retention of motor skill after motor imagery training. *Neuropsychologia*, 143, 107472. <https://doi.org/10.1016/j.neuropsychologia.2020.107472>
- *Bonczyk, C. S., Schroeder, K. M., Anderson, B., & Galgon, R. E. (2016). Two methods for teaching basic upper airway sonography. *Journal of Clinical Anesthesia*, 31, 166-172. <https://doi.org/10.1016/j.jclinane.2016.01.040>

- *Bonuzzi, G. M. G., Alves, É. J. M., & Perotti, A. (2020). Effects of the aerobic exercise on the learning of a sports motor skill. *Motriz. Revista de Educacao Fisica*, 26(2), e10201420. <https://doi.org/10.1590/s1980-6574202000011420>
- *Bootsma, J. M., Hortobágyi, T., Rothwell, J. C., & Caljouw, S. R. (2018). The role of task difficulty in learning a visuomotor skill. *Medicine and Science in Sports and Exercise*, 50(9), 1842-1849. <https://doi.org/10.1249/mss.0000000000001635>
- Borenstein, M. Cooper, H., Hedges, L., & Valentine, J. (2009). Effect sizes for continuous data. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd, pp. 221–235). New York: Sage.
- *Bouchard, L. J., & Singer, R. N. (1998). Effects of the five-step strategy with videotape modeling on performance of the tennis serve. *Perceptual and Motor Skills*, 86(3 Pt 1), 739-746. <https://doi.org/10.2466/pms.1998.86.3.739>
- *Bouwsema, H., van der Sluis, C. K., & Bongers, R. M. (2014). Changes in performance over time while learning to use a myoelectric prosthesis. *Journal of Neuroengineering and Rehabilitation* 11, 16. <https://doi.org/10.1186/1743-0003-11-16>
- *Boyce, B. A. (1990). The effect of instructor-set goals upon skill acquisition and retention of a selected shooting task. *Journal of Teaching in Physical Education*, 9(2), 115-122. <https://doi.org/10.1123/jtpe.9.2.115>
- *Boyce, B. A. (1992). The effects of three styles of teaching on university students' motor performance. *Journal of Teaching in Physical Education*, 11(4), 389-401. <https://doi.org/10.1123/jtpe.11.4.389>
- Braun, L., Sawyer, T., Kavanagh, L., & Deering, S. (2014). Facilitating physician reentry to practice: perceived effects of deployments on US army pediatricians' clinical and procedural skills. *Journal of Continuing Education in the Health Professions*, 34(4), 252-259. <https://doi.org/10.1002/chp.21256>
- *Brawn, T. P., Fenn, K. M., Nusbaum, H. C., & Margoliash, D. (2008). Consolidation of sensorimotor learning during sleep. *Learning and Memory*, 15(11), 815-819. <https://doi.org/10.1101/lm.1180908>

- *Breslin, G., Hodges, N. J., Williams, A. M., Curran, W., & Kremer, J. (2005). Modelling relative motion to facilitate intra-limb coordination. *Human Movement Science*, 24(3), 446-463. <https://doi.org/10.1016/j.humov.2005.06.009>
- Brevig, H., Colosimo, C., Jaditz, T., Krauss, R., Mandell, K., Morrow, R., Oi, J., & Tsang, W. (2015). The quality-volume relationship: Comparing civilian and MHS practice. *Center for Naval Analyses*. Alexandria, VA: CNA Analysis and Solutions. <https://www.cna.org/reports/2014/DIM-2014-U-009221-Final.pdf>
- *Broadbent, D. P., Ford, P. R., O'Hara, D. A., Williams, A. M., & Causer, J. (2017). The effect of a sequential structure of practice for the training of perceptual-cognitive skills in tennis. *PLoS ONE*, 12(3), 0174311. <https://doi.org/10.1371/journal.pone.0174311>
- Brown, T. L., & Carr, T. H. (1989). Automaticity in skill acquisition: Mechanisms for reducing interference in concurrent performance. *Journal of Experimental Psychology: Human Perception and Performance*, 15(4), 686-700. <https://doi.org/10.1037/0096-1523.15.4.686>
- *Brueckner, D., Göpfert, B., Kiss, R., & Muehlbauer, T. (2019). Effects of motor practice on learning a dynamic balance task in healthy young adults: A wavelet-based time-frequency analysis. *Gait and Posture*, 70, 264-269. <https://doi.org/10.1016/j.gaitpost.2019.03.019>
- *Bruppacher, H. R., Alam, S. K., LeBlanc, V. R., Latter, D., Naik, V. N., Savoldelli, G. L., Mazer, C. D., Kurrek, M. M., & Joo, H. S. (2010). Simulation-based training improves physicians' performance in patient care in high-stakes clinical setting of cardiac surgery. *Anesthesiology*, 112(4), 985-992. <https://doi.org/10.1097/ALN.0b013e3181d3e31c>
- Bryant, D. J. & Angel, H. (2000). *Retention and fading of military skills: Literature review* (DCIEM No. CR 2000-07). Guelph, ON: Human Systems Incorporated. <https://apps.dtic.mil/sti/citations/ADA593268>
- *Buchanan, J. J., & Wright, D. L. (2011). Generalization of action knowledge following observational learning. *Acta Psychologica*, 136(1), 167-178. <https://doi.org/10.1016/j.actpsy.2010.11.006>
- *Buescher, J. F., Mehdorn, A. S., Neumann, P. A., Becker, F., Eichelmann, A. K., Pankratius, U., Bahde, R., Foell, D., Senninger, N., & Rijcken, E. (2018). Effect of continuous motion parameter feedback on laparoscopic simulation training: A

prospective randomized controlled trial on skill acquisition and retention. *Journal of Surgical Education*, 75(2), 516-526. <https://doi.org/10.1016/j.jsurg.2017.08.015>

- *Bullard, S., & Miller, S. E. (2001). Comparison of teaching methods to learn a tilt and balance wheelchair skill. *Perceptual and Motor Skills*, 93(1), 131-138. <https://doi.org/10.2466/pms.2001.93.1.131>

- *Burns, G. T., King, B. W., Holmes, J. R., & Irwin, T. A. (2017). Evaluating internal fixation skills using surgical simulation. *Journal of Bone and Joint Surgery - American Volume*, 99(5), e21.21-e21.28. <https://doi.org/10.2106/JBJS.16.00653>

- *Carlsen, C. G., Lindorff-Larsen, K., Funch-Jensen, P., Lund, L., Konge, L., & Charles, P. (2015). Module based training improves and sustains surgical skills: a randomised controlled trial. *Hernia*, 19(5), 755-763. <https://doi.org/10.1007/s10029-015-1357-6>

- *Carter, M. J., Carlsen, A. N., & Ste-Marie, D. M. (2014). Self-controlled feedback is effective if it is based on the learner's performance: A replication and extension of Chiviawosky and Wulf (2005). *Frontiers in Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.01325>

- *Carter, M. J., Smith, V., & Ste-Marie, D. M. (2016). Judgments of learning are significantly higher following feedback on relatively good versus relatively poor trials despite no actual learning differences. *Human Movement Science*, 45, 63-70. <https://doi.org/10.1016/j.humov.2015.11.006>

- *Carter, M. J., & Ste-Marie, D. M. (2017). An interpolated activity during the knowledge-of-results delay interval eliminates the learning advantages of self-controlled feedback schedules. *Psychological Research*, 81(2), 399-406. <https://doi.org/10.1007/s00426-016-0757-2>

- *Casabona, A., Lombardo, L., Cioni, M., & Valle, M. S. (2018). Delayed benefits from spaced training when learning a precision throwing task. *Applied Sciences*, 8(12), 2359. <https://doi.org/10.3390/app8122359>

- *Casabona, A., Valle, M. S., Cavallaro, C., Castorina, G., & Cioni, M. (2018). Selective improvements in balancing associated with offline periods of spaced training. *Scientific Reports*, 8, 7836. <https://doi.org/10.1038/s41598-018-26228-4>

- *Castellvi, A. O., Hollett, L. A., Minhajuddin, A., Hogg, D. C., Tesfay, S. T., & Scott, D. J. (2009). Maintaining proficiency after fundamentals of laparoscopic surgery training: a 1-year analysis of skill retention for surgery residents. *Surgery, 146*(2), 387-393. <https://doi.org/10.1016/j.surg.2009.05.009>
- *Chamberlin, C. J., & Magill, R. A. (1992). The memory representation of motor skills: A test of schema theory. *Journal of Motor Behavior, 24*(4), 309-319. <https://doi.org/10.1080/00222895.1992.9941627>
- *Chan, A., Singh, S., Dubrowski, A., Pratt, D. D., Zalunardo, N., Nair, P., McLaughlin, K., & Ma, I. W. Y. (2015). Part versus whole: A randomized trial of central venous catheterization education. *Advances in Health Sciences Education, 20*(4), 1061-1071. <https://doi.org/10.1007/s10459-015-9586-0>
- *Chartrand, G., Kaneva, P., Kolozsvari, N., Li, C., Petrucci, A. M., Mutter, A. F., Daskalopoulou, S. S., Carli, F., Feldman, L. S., Fried, G. M., & Vassiliou, M. C. (2015). The effects of acute aerobic exercise on the acquisition and retention of laparoscopic skills. *Surgical Endoscopy, 29*(2), 474-480. <https://doi.org/10.1007/s00464-014-3691-7>
- *Chauvel, G., Wulf, G., & Maquestiaux, F. (2015). Visual illusions can facilitate sport skill learning. *Psychonomic Bulletin & Review, 22*(3), 717-721. <https://doi.org/10.3758/s13423-014-0744-9>
- *Cheung, J. J. (2014). *Preparing for simulation-based education and training through web-based learning: The role of observational practice and educational networking* (Publication No. 1033015213). [Master's thesis, University of Toronto]. <https://tspace.library.utoronto.ca/handle/1807/44009>
- *Cheung, J. J. (2019). *Knowing how and knowing why: Integrating conceptual knowledge in simulation-based procedural skills training to support learning transfer* (Publication No. 22624351). [Doctoral dissertation, University of Toronto]. ProQuest Dissertations & Theses Global.
- Chi, M. T. H., & Ohlsson, S. 2005. Complex declarative learning. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 371-399). New York: Cambridge University Press.
- *Chien, C. Y., Fang, S. Y., Tsai, L. H., Tsai, S. L., Chen, C. B., Seak, C. J., Weng, Y. M., Lin, C. C., Chien, W. C., Huang, C. H., Lin, C. Y., Chaou, C. H., Liu, P. H., Tseng, H. J., Chen, J. C., Peng, S. Y., Cheng, T. H., Hsu, K. H., & Ng, C. J.

- (2020). Traditional versus blended CPR training program: A randomized controlled non-inferiority study. *Scientific Reports*, 10(1), 10032. <https://doi.org/10.1038/s41598-020-67193-1>
- *Choi, H. J., Lee, C. C., Lim, T. H., Kang, B. S., Singer, A. J., & Henry, M. C. (2010). Effectiveness of mouth-to-mouth ventilation after video self-instruction training in laypersons. *American Journal of Emergency Medicine*, 28(6), 654-657. <https://doi.org/10.1016/j.ajem.2009.02.015>
- *Christova, M., Rafolt, D., & Gallasch, E. (2015). Cumulative effects of anodal and priming cathodal tDCS on pegboard test performance and motor cortical excitability. *Behavioural Brain Research*, 287, 27-33. <https://doi.org/10.1016/j.bbr.2015.03.028>
- *Chiviawosky, S., Harter, N. M., Gonçalves, G. S., & Cardozo, P. L. (2019). Temporal-comparative feedback facilitates golf putting. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.02691>
- *Chu, D. P. K. (2006). The effects of augmented feedback training in cadence acquisition. *Research in Sports Medicine*, 14(2), 135-147. <https://doi.org/10.1080/15438620600651165>
- *Ciechanski, P., Cheng, A., Damji, O., Lopushinsky, S., Hecker, K., Jadavji, Z., & Kirton, A. (2018). Effects of transcranial direct-current stimulation on laparoscopic surgical skill acquisition. *BJS Open*, 2(2), 70-78. <https://doi.org/10.1002/bjs5.43>
- *Clark, C. A., Mester, R. A., Redding, A. T., Wilson, D. A., Zeiler, L. L., Jones, W. R., Reves, J. G., Reeves, S. T., & Schaefer, J. J. (2022). Emergency subglottic airway training and assessment of skills retention of attending anesthesiologists with simulation mastery-based learning. *Anesthesia and Analgesia*, 135(1), 143-151. <https://doi.org/10.1213/ANE.0000000000005928>
- *Clinkard, D., Moulton, E., Holden, M., Davison, C., Ungi, T., Fichtinger, G., & McGraw, R. (2015). Assessment of lumbar puncture skill in experts and nonexperts using checklists and quantitative tracking of needle trajectories: Implications for competency-based medical education. *Teaching and Learning in Medicine*, 27(1), 51-56. <https://doi.org/10.1080/10401334.2014.979184>
- *Collins, A. M., Ridgway, P. F., Hassan, M. S., Chou, C. W., Hill, A. D., & Kneafsey, B. (2010). Surgical instruction for general practitioners: how, who and how often?

Journal of Plastic, Reconstructive & Aesthetic Surgery, 63(7), 1156-1162.
<https://doi.org/10.1016/j.bjps.2009.05.023>

- *Compton, N. J., Cary, J. A., Wenz, J. R., Lutter, J. D., Mitchell, C. F., & Godfrey, J. (2019). Evaluation of peer teaching and deliberate practice to teach veterinary surgery. *Veterinary Surgery*, 48(2), 199-208. <https://doi.org/10.1111/vsu.13117>

- *Conroy, S. M., Bond, W. F., Pheasant, K. S., & Ceccacci, N. (2010). Competence and retention in performance of the lumbar puncture procedure in a task trainer model. *Simulation in Healthcare*, 5(3), 133-138.
<https://doi.org/10.1097/SIH.0b013e3181dc040a>

- *Corbin, C. B. (1966). *The effects of mental practice on the development of a specific motor skill* (Publication No. 6603092). [Doctoral dissertation, University of New Mexico]. ProQuest Dissertations & Theses Global.

- *Coughlan, E. K., Williams, A. M., McRobert, A. P., & Ford, P. R. (2014). How experts practice: A novel test of deliberate practice theory. *Journal of Experimental Psychology: Learning Memory and Cognition*, 40(2), 449-458.
<https://doi.org/10.1037/a0034302>

- *Couvillion, K. F., Bass, A. D., & Fairbrother, J. T. (2020). Increased cognitive load during acquisition of a continuous task eliminates the learning effects of self-controlled knowledge of results. *Journal of Sports Sciences*, 38(1), 94-99.
<https://doi.org/10.1080/02640414.2019.1682901>

- *Coxon, J. P., Peat, N. M., & Byblow, W. D. (2014). Primary motor cortex disinhibition during motor skill learning. *Journal of Neurophysiology*, 112(1), 156-164.
<https://doi.org/10.1152/jn.00893.2013>

- *Crewther, B. T., Shetty, K., Jarchi, D., Selvadurai, S., Cook, C. J., Leff, D. R., Darzi, A., & Yang, G. Z. (2016). Skill acquisition and stress adaptations following laparoscopic surgery training and detraining in novice surgeons. *Surgical Endoscopy*, 30(7), 2961-2968. <https://doi.org/10.1007/s00464-015-4584-0>

- *Cuppone, A. V., Semprini, M., & Konczak, J. (2018). Consolidation of human somatosensory memory during motor learning. *Behavioural Brain Research*, 347, 184-192. <https://doi.org/10.1016/j.bbr.2018.03.013>

- *Curran, V. R., Aziz, K., O'Young, S., & Bessell, C. (2004). Evaluation of the effect of a computerized training simulator (ANAKIN) on the retention of neonatal resuscitation skills. *Teaching and Learning in Medicine, 16*(2), 157-164. https://doi.org/10.1207/s15328015tlm1602_7
- *Currat, L., Suppan, M., Gartner, B. A., Daniel, E., Mayoraz, M., Harbarth, S., Suppan, L., & Stuby, L. (2022). Impact of face-to-face teaching in addition to electronic learning on personal protective equipment doffing proficiency in student paramedics: randomized controlled trial. *International Journal of Environmental Research and Public Health, 19*(5), 3077. <https://doi.org/10.3390/ijerph19053077>
- Dalton, D. R., Aguinis, H., Dalton, C. M., Bosco, F. A., & Pierce, C. A. (2012). Revisiting the file drawer problem in meta-analysis: An assessment of published and nonpublished correlation matrices. *Personnel Psychology, 65*(2), 221-249. <https://doi.org/10.1111/j.1744-6570.2012.01243.x>
- *Dancey, E., Murphy, B. A., Andrew, D., & Yelder, P. (2016). The effect of local vs remote experimental pain on motor learning and sensorimotor integration using a complex typing task. *Pain, 157*(8), 1682-1695. <https://doi.org/10.1097/j.pain.0000000000000570>
- D'Angelo, A. L. D., D'Angelo, J. D., Rogers, D. A., & Pugh, C. M. (2018). Faculty perceptions of resident skills decay during dedicated research fellowships. *American Journal of Surgery, 215*(2), 336-340. <https://doi.org/10.1016/j.amjsurg.2017.11.018>
- *Dantas, A. K., Shinagawa, A., & Deboni, M. C. (2010). Assessment of preclinical learning on oral surgery using three instructional strategies. *Journal of Dental Education, 74*(11), 1230-1236. <https://doi.org/10.1002/j.0022-0337.2010.74.11.tb04997.x>
- *Dayan, E., Averbeck, B. B., Richmond, B. J., & Cohen, L. G. (2014). Stochastic reinforcement benefits skill acquisition. *Learning & Memory, 21*(3), 140-142. <https://doi.org/10.1101/lm.032417.113>
- *Delasobera, B. E., Goodwin, T. L., Strehlow, M., Gilbert, G., D'Souza, P., Alok, A., Raje, P., & Mahadevan, S. V. (2010). Evaluating the efficacy of simulators and multimedia for refreshing ACLS skills in India. *Resuscitation, 81*(2), 217-223. <https://doi.org/10.1016/j.resuscitation.2009.10.013>

- *De Groot, S., De Winter, J. C. F., García, J. M. L., Mulder, M., & Wieringa, P. A. (2011). The effect of concurrent bandwidth feedback on learning the lane-keeping task in a driving simulator. *Human Factors*, 53(1), 50-62.
<https://doi.org/10.1177/0018720810393241>
- *Demirtas, A., Basak, T., Sahin, G., & Sonkaya, M. Ç. (2022). The serious game and integrated simulator for cardiopulmonary resuscitation training in nursing students. *Simulation & Gaming*, 53(2), 97-110.
<https://doi.org/10.1177/10468781211073162>
- *De Regge, M., Calle, P. A., De Paepe, P., & Monsieurs, K. G. (2008). Basic life support refresher training of nurses: individual training and group training are equally effective. *Resuscitation*, 79(2), 283-287.
<https://doi.org/10.1016/j.resuscitation.2008.06.015>
- *Deschamps, T., Nourrit, D., Caillou, N., & Delignières, D. (2004). Influence of a stressing constraint on stiffness and damping functions of a ski simulator's platform motion. *Journal of Sports Sciences*, 22(9), 867-874.
<https://doi.org/10.1080/02640410410001675478>
- *Devine, L. A., Donkers, J., Brydges, R., Perelman, V., Cavalcanti, R. B., & Issenberg, S. B. (2015). An equivalence trial comparing instructor-regulated with directed self-regulated mastery learning of advanced cardiac life support skills. *Simulation in Healthcare*, 10(4), 202-209. <https://doi.org/10.1097/sih.0000000000000095>
- *Diederich, E., Thomas, L., Mahnken, J., & Lineberry, M. (2018). Pretest Scores Uniquely Predict 1-Year-Delayed Performance in a Simulation-Based Mastery Course for Central Line Insertion. *Simulation in Healthcare*, 13(3), 163-167.
<https://doi.org/10.1097/SIH.0000000000000327>
- *Drews, R., Pacheco, M. M., Bastos, F. H., & Tani, G. (2021). Knowledge of results do not affect self-efficacy and skill acquisition on an anticipatory timing task. *Journal of Motor Behavior*, 53(3), 275-286.
<https://doi.org/10.1080/00222895.2020.1772711>
- Driskell, J. E., Willis, R. P., & Copper, C. (1992). Effect of overlearning on retention. *Journal of Applied Psychology*, 77(5), 615–622.
- Duffin, E. (2022). Armed forces of the United States – Statistics & Facts. *Statista*. Retrieved from <https://www.statista.com/topics/2171/armed-forces-of-the-united-states/#topicOverview>

- Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched group or repeated measures designs. *Psychological Methods, 1*(2), 170-177. <https://doi.org/10.1037/1082-989X.1.2.170>
- *Dunton, A., O'Neill, C., & Coughlan, E. K. (2020). The impact of a spatial occlusion training intervention on pass accuracy across a continuum of representative experimental design in football. *Science and Medicine in Football, 4*(4), 269-277. <https://doi.org/10.1080/24733938.2020.1745263>
- Durning, S. J., Artino, A. R., Holmboe, E., Beckman, T. J., Van Der Vleuten, C., & Schuwirth, L. (2010). Aging and cognitive performance: challenges and implications for physicians practicing in the 21st century. *Journal of Continuing Education in the Health Professions, 30*(3), 153-160. <https://doi.org/10.1002/chp.20075>
- *Dyer, J. F., Stapleton, P., & Rodger, M. W. M. (2017). Transposing musical skill: sonification of movement as concurrent augmented feedback enhances learning in a bimanual task. *Psychological Research, 81*(4), 850-862. <https://doi.org/10.1007/s00426-016-0775-0>
- *Dyer, J. F., Stapleton, P., & Rodger, M. W. M. (2017). Advantages of melodic over rhythmic movement sonification in bimanual motor skill learning. *Experimental Brain Research, 235*(10), 3129-3140. <https://doi.org/10.1007/s00221-017-5047-8>
- *van Dongen, K. W., Mitra, P. J., Schijven, M. P., & Broeders, I. (2011). Distributed versus massed training: efficiency of training psychomotor skills. *Surgical Techniques Development, 1*(1), 40-42, e17. <https://doi.org/10.4081/std.2011.e17>
- *Echt, K. V., Morrell, R. W., & Park, D. C. (1998). Effects of age and training formats on basic computer skill acquisition in older adults. *Educational Gerontology, 24*(1), 3-25. <https://doi.org/10.1080/0360127980240101>
- *Edelman, D. A., Mattos, M. A., & Bouwman, D. L. (2010). FLS skill retention (learning) in first year surgery residents. *Journal of Surgical Research, 163*(1), 24-28. <https://doi.org/10.1016/j.jss.2010.03.057>
- *Eley, J. G., & Birnie, C. (2006). Retention of compounding skills among pharmacy students. *American Journal of Pharmaceutical Education, 70*(6), 132. <https://doi.org/10.5688/aj7006132>

- *Elion, O., Sela, I., Bahat, Y., Siev-Ner, I., Weiss, P. L., & Karni, A. (2015). Balance maintenance as an acquired motor skill: Delayed gains and robust retention after a single session of training in a virtual environment. *Brain Research, 1609*, 54-62. <https://doi.org/10.1016/j.brainres.2015.03.020>
- *Ellis, J. A., Whitehill, B. V., & Irick, C. (1996). The effects of explanations and pictures on learning, retention, and transfer of a procedural assembly task. *Contemporary Educational Psychology, 21*(2), 129-148. <https://doi.org/10.1006/ceps.1996.0012>
- *Ellison, P., Jones, C., Sparks, S. A., Murphy, P. N., Page, R. M., Carnegie, E., & Marchant, D. C. (2020). The effect of stroboscopic visual training on eye-hand coordination. *Sport Sciences for Health, 16*(3), 401-410. <https://doi.org/10.1007/s11332-019-00615-4>
- *Etnier, J. L., & Landers, D. M. (1998). Motor performance and motor learning as a function of age and fitness. *Research Quarterly for Exercise and Sport, 69*(2), 136-146. <https://doi.org/10.1080/02701367.1998.10607679>
- *Etnier, J. L., Romero, D. H., & Traustadóttir, T. (2001). Acquisition and Retention of Motor Skills as a Function of Age and Aerobic Fitness. *Journal of Aging and Physical Activity, 9*(4), 425-437. <https://doi.org/10.1123/japa.9.4.425>
- *Etnyre, B. R., & Poindexter, H. B. (1995). Characteristics of motor performance, learning, warm-up decrement, and reminiscence during a balancing task. *Perceptual and Motor Skills, 80*(3), 1027-1030. <https://doi.org/10.2466/pms.1995.80.3.1027>
- *Ewolds, H., Broeker, L., De Oliveira, R. F., Raab, M., & KÜNZell, S. (2020). Ways to improve multitasking: Effects of predictability after single- And dual-task training. *Journal of Cognition, 4*(1), A16. <https://doi.org/10.5334/joc.142>
- *Eysenck, S. B. G. (1960). Retention of a well-developed motor skill after one year. *Journal of General Psychology, 63*, 267-273. <https://doi.org/10.1080/00221309.1960.9711824>
- Farr, M. J. (1987). *The long-term retention of knowledge and skills: A cognitive and instructional perspective*. NY: Springer-Verlag.

- *Fazeli, D., Taheri, H., & Saberi Kakhki, A. (2017). Random versus blocked practice to enhance mental representation in golf putting. *Perceptual and Motor Skills*, 124(3), 674-688. <https://doi.org/10.1177/0031512517704106>
- Fendrich D. W, Healy, A. F., Meiskey, L., Crutcher, R. J., Little, W., Bourne Jr., L. E. (1988). *Skill maintenance: Literature review and theoretical analysis* (AFHRL-TP-87-73). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.
- *Feng, D. B., Yong, Y. H., Byrnes, T., Gorelik, A., Colt, H., Irving, L. B., & Steinfert, D. (2020). Learning gain and skill retention following unstructured bronchoscopy simulation in a low-fidelity airway model. *Journal of Bronchology & Interventional Pulmonology*, 27(4), 280-285. <https://doi.org/10.1097/lbr.0000000000000664>
- Ferguson, C. J., & Brannick, M. T. (2012). Publication bias in psychological science: Prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. *Psychological Methods*, 17(1), 120–128. <https://doi.org/10.1037/a0024445>
- *Feygin, D. (2002). *Implementation and validation of haptic training in virtual surgical environments* (Publication No. 3082182). [Doctoral dissertation, University of California, Berkeley]. ProQuest Dissertations & Theses Global.
- *Fisher, K. M., & Fairbrother, J. T. (2020). Seeing is believing: Blind putting drills confer no advantage to the novice golfer. *Research Quarterly for Exercise and Sport*, 91(2), 335-345. <https://doi.org/10.1080/02701367.2019.1674443>
- Fitts, P. M., & Posner, M. I. (1967). *Human performance*. Belmont, CA: Brooks/Cole Publishing.
- Fleishman, E. A. (1954). Dimensional analysis of psychomotor abilities. *Journal of Experimental Psychology*, 48(6), 437–454. <https://doi.org/10.1037/h0058244>
- *Fleishman, E. A., & Parker, J. F., Jr. (1962). Factors in the retention and relearning of perceptual-motor skill. *Journal of Experimental Psychology*, 64(3), 215–226. <https://doi.org/10.1037/h0041220>
- *Fotheringham, J. (1984). Transfer of training: A field investigation of youth training. *Journal of Occupational Psychology*, 57(3), 239-248. <https://doi.org/10.1111/j.2044-8325.1984.tb00165.x>

- *Frank, C., Kim, T., & Schack, T. (2018). Observational practice promotes action-related order formation in long-term memory: Investigating action observation and the development of cognitive representation in complex motor action. *Journal of Motor Learning and Development*, 6(1), 53-72. <https://doi.org/10.1123/jmld.2017-0007>
- *Friederichs, H., Marschall, B., & Weissenstein, A. (2019). Simulation-based mastery learning in medical students: Skill retention at 1-year follow up. *Medical Teacher*, 41(5), 539-546. <https://doi.org/10.1080/0142159X.2018.1503411>
- *Frikha, M., Chaâri, N., Elghoul, Y., Mohamed-Ali, H. H., & Zinkovsky, A. V. (2019). Effects of combined versus singular verbal or haptic feedback on acquisition, retention, difficulty, and competence perceptions in motor learning. *Perceptual and Motor Skills*, 126(4), 713-732. <https://doi.org/10.1177/0031512519842759>
- *Frithioff, A., Frendo, M., Von Buchwald, J. H., Mikkelsen, P. T., Sorensen, M. S., & Andersen, S. A. W. (2022). Automated summative feedback improves performance and retention in simulation training of mastoidectomy: a randomised controlled trial. *Journal of Laryngology and Otology*, 136(1), 29-36. <https://doi.org/10.1017/S0022215121003352>
- *Frömer, R., Stürmer, B., & Sommer, W. (2016). (Don't) Mind the effort: Effects of contextual interference on ERP indicators of motor preparation. *Psychophysiology*, 53(10), 1577-1586. <https://doi.org/10.1111/psyp.12703>
- *Gabitov, E., Boutin, A., Pinsard, B., Censor, N., Fogel, S. M., Albouy, G., King, B. R., Carrier, J., Cohen, L. G., Karni, A., & Doyon, J. (2019). Susceptibility of consolidated procedural memory to interference is independent of its active task-based retrieval. *PLoS ONE*, 14(1), e0210876. <https://doi.org/10.1371/journal.pone.0210876>
- *Gabriele, K. M., Holthaus, R. M., & Boulet, J. R. (2016). Usefulness of video-assisted peer mentor feedback in undergraduate nursing education. *Clinical Simulation in Nursing*, 12(8), 337-345. <https://doi.org/10.1016/j.ecns.2016.03.004>
- *Gabriele, T. E., Hall, C. R., & Lee, T. D. (1989). Cognition in motor learning: Imagery effects on contextual interference. *Human Movement Science*, 8(3), 227-245. [https://doi.org/10.1016/0167-9457\(89\)90008-0](https://doi.org/10.1016/0167-9457(89)90008-0)
- *Gal, C., Gabitov, E., Maaravi-Hesseg, R., Karni, A., & Korman, M. (2019). A delayed advantage: Multi-session training at evening hours Leads to better long-term

retention of motor skill in the elderly. *Frontiers in Aging Neuroscience*, 11, 321.
<https://doi.org/10.3389/fnagi.2019.00321>

Gardlin, G. R., & Sitterly, T. E. (1972). *Degradation of learned skills: A review and annotated bibliography* (Document No. D180-15080-1). Seattle, WA: The Boeing Co. <https://ntrs.nasa.gov/citations/19730001425>

*Gardner, A. K., Abdelfattah, K., Wiersch, J., Ahmed, R. A., & Willis, R. E. (2015). Embracing errors in simulation-based training: The effect of error training on retention and transfer of central venous catheter skills. *Journal of Surgical Education*, 72(6), e158-162. <https://doi.org/10.1016/j.jsurg.2015.08.002>

*Gerard, J. M., Thomas, S. M., Germino, K. W., Street, M. H., Burch, W., & Scalzo, A. J. (2011). The effect of simulation training on PALS skills among family medicine residents. *Family Medicine*, 43(6), 392-399.

*Ghorbani, S. (2019). Motivational effects of enhancing expectancies and autonomy for motor learning: An examination of the OPTIMAL theory. *Journal of General Psychology*, 146(1), 79-92. <https://doi.org/10.1080/00221309.2018.1535486>

*Ghotbaldinian, E., Dehdari, N., & Åkeson, J. (2021). Maintenance of basic endotracheal intubation skills with direct or video-assisted laryngoscopy: A randomized crossover follow-up study in inexperienced operators. *AEM Education and Training*, 5(4), e10655. <https://doi.org/10.1002/aet2.10655>

Gibb, G. D., & Dolgin, D. L. (1989). Predicting military flight training success by a compensatory tracking task. *Military Psychology*, 1(4), 235-240.
https://doi.org/10.1207/s15327876mp0104_4

Gillett, B., Saloum, D., Aghera, A., & Marshall, J. P. (2019). Skill proficiency is predicted by intubation frequency of emergency medicine attending physicians. *Western Journal of Emergency Medicine*, 20(4), 601- 609.
<https://doi.org/10.5811%2Fwestjem.2019.6.42946>

*van Ginneken, W. F., Poolton, J. M., Capio, C. M., van der Kamp, J., Choi, C. S. Y., & Masters, R. S. W. (2018). Conscious control is associated with freezing of mechanical degrees of freedom during motor learning. *Journal of Motor Behavior*, 50(4), 436-456. <https://doi.org/10.1080/00222895.2017.1365045>

- *Goedert, K. M., & Miller, J. (2008). Spacing practice sessions across days earlier rather than later in training improves performance of a visuomotor skill. *Experimental Brain Research*, 189(2), 189-197. <https://doi.org/10.1007/s00221-008-1414-9>
- *Gofton, W., Dubrowski, A., Tabloie, F., & Backstein, D. (2007). The effect of computer navigation on trainee learning of surgical skills. *Journal of Bone and Joint Surgery*, 89(12), 2819-2827. <https://doi.org/10.2106/jbjs.F.01502>
- *Goldman, M. F. (1972). *The learning, retention, and bilateral transfer of a motor skill by college women as a function of mental practice, physical practice and mixed practice* (Publication No. 7226595). [Doctoral dissertation, New York University]. ProQuest Dissertations & Theses Global
- *Gomes, T. V. B., Ugrinowitsch, H., Marinho, N., Shea, J. B., Raisbeck, L. D., & Benda, R. N. (2014). Effects of mental practice in novice learners in a serial positioning skill acquisition. *Perceptual and Motor Skills*, 119(2), 397-414. <https://doi.org/10.2466/23.PMS.119c20z4>
- *Goodwin, J. E., Grimes, C. R., Eckerson, J. M., & Gordon, P. M. (1998). Effect of different quantities of variable practice on acquisition, retention, and transfer of an applied motor skill. *Perceptual and Motor Skills*, 87(1), 147-151. <https://doi.org/10.2466/pms.1998.87.1.147>
- *Granados, C., & Wulf, G. (2007). Enhancing motor learning through dyad practice: Contributions of observation and dialogue. *Research Quarterly for Exercise and Sport*, 78(3), 197-203. <https://doi.org/10.1080/02701367.2007.10599417>
- *Grierson, L. E., Lyons, J. L., & Dubrowski, A. (2013). Gaze-down endoscopic practise leads to better novice performance on gaze-up displays. *Med Education*, 47(2), 166-172. <https://doi.org/10.1111/medu.12084>
- *Green, L. A., McGuire, J., & Gabriel, D. A. (2021). Effects of the proprioceptive neuromuscular facilitation contraction sequence on motor skill learning-related increases in the maximal rate of wrist flexion torque development. *Frontiers in Human Neuroscience*, 15, 764660. <https://doi.org/10.3389/fnhum.2021.764660>
- *Grimsley, D. L. (1969). *Acquisition, retention, and retraining: Effects of high and low fidelity in training devices* (Technical Report 69-1). Alexandria, VA: Human Resources Research Office. <https://apps.dtic.mil/sti/citations/AD0685074>

- *Gurung, P. M., Campbell, T., Wang, B., Joseph, J. V., & Ghazi, A. E. (2020). Accelerated Skills Acquisition Protocol (ASAP) in optimizing robotic surgical simulation training: a prospective randomized study. *World Journal of Urology*, 38, 1623-1630. <https://doi.org/10.1007/s00345-019-02858-9>
- *Habibli, T., Ghezeljeh, T. N., & Haghani, S. (2020). The effect of simulation-based education on nursing students' knowledge and performance of adult basic cardiopulmonary resuscitation: A randomized clinical trial. *Nursing Practice Today*, 7(2), 87-96.
- *Hagedorn, D. W. (1994). *The Effects of Part-Training on Cardiopulmonary Resuscitation Skill Acquisition and Retention*. [Master's thesis, Western Michigan University]. WMU University Libraries. https://scholarworks.wmich.edu/masters_theses/4640/
- Hagman, J. D., & Rose, A. M. (1983). Retention of military tasks: A review. *Human Factors*, 25(2), 199-213. <https://doi.org/10.1177/001872088302500207>
- *Haji, F. A., Cheung, J. J. H., Woods, N., Regehr, G., de Ribaupierre, S., & Dubrowski, A. (2016). Thrive or overload? The effect of task complexity on novices' simulation-based learning. *Medical Education*, 50(9), 955-968. <https://doi.org/10.1111/medu.13086>
- *Hamann, J. M., Dayan, E., Hummel, F. C., & Cohen, L. G. (2014). Baseline frontostriatal-limbic connectivity predicts reward-based memory formation. *Human Brain Mapping*, 35(12), 5921-5931. <https://doi.org/10.1002/hbm.22594>
- *Handley, A. J. (2002). Teaching hand placement for chest compression--a simpler technique. *Resuscitation*, 53(1), 29-36. [https://doi.org/10.1016/s0300-9572\(01\)00506-8](https://doi.org/10.1016/s0300-9572(01)00506-8)
- *Hartman, J. M. (2007). Self-controlled use of a perceived physical assistance device during a balancing task. *Perceptual and Motor Skills*, 104(3, Pt 1), 1005-1016. <https://doi.org/10.2466/PMS.104.3.1005-1016>
- *Hasdianda, M. A. (2016). *Understanding the impact of acute stressor during simulation on medical students' short and long-term clinical skills retention* (Publication No. 28478295). [Master's thesis, Harvard University]. ProQuest Dissertations & Theses Global.

- *Hautala, R. M., & Conn, J. H. (1993). A test of Magill's closed-to-open continuum for skill development. *Perceptual and Motor Skills*, 77(1), 219-226.
<https://doi.org/10.2466/pms.1993.77.1.219>
- Haworth, N., Mulvhill, C., & Symmons, M. (2002). *Motorcycling after 30* (Report #192). Clayton, VIC: Monash University Accident Research Center. Monash University Archives. <https://www.monash.edu/muarc/archive/our-publications/reports/muarc192>
- *Hayes, S. J., Hodges, N. J., Huys, R., & Williams, A. M. (2007). End-point focus manipulations to determine what information is used during observational learning. *Acta Psychologica*, 126(2), 120-137.
<https://doi.org/10.1016/j.actpsy.2006.11.003>
- *Healy, A. F., Kole, J. A., Schneider, V. I., & Barshi, I. (2019). Training, retention, and transfer of data entry perceptual and motor processes over short and long retention intervals. *Memory and Cognition*, 47(8), 1606-1618.
<https://doi.org/10.3758/s13421-019-00955-z>
- *Heitman, R. J., Pugh, S. F., Kovalski, J. E., Norell, P. M., & Vicory, J. R. (2005). Effects of specific versus variable practice on the retention and transfer of a continuous motor skill. *Perceptual and Motor Skills*, 100(3 II), 1107-1113.
<https://doi.org/10.2466/pms.100.3c.1107-1113>
- *Hempel, G., Heinke, W., Struck, M. F., Piegeler, T., & Rotzoll, D. (2019). Impact of quantitative feedback via high-fidelity airway management training on success rate in endotracheal intubation in undergraduate medical students-A prospective single-center study. *Journal of Clinical Medicine*, 8(9), 1465.
<https://doi.org/10.3390/jcm8091465>
- Henik, A., Brain, E., Ze'evi, V., & Schwartz, D. (1999). *The preservation and the decay of military skills* (R&D 7699-RB-01). Israel: The Psychology Branch Research Section IDF. <https://apps.dtic.mil/sti/citations/ADA389399>
- *Hernandez, R. (2017). *Effect of visual feedback on the attainment and retention of force moderation in laparoscopic Training* (Publication No. 10283901). [Doctoral dissertation, University of Miami]. ProQuest Dissertations and Theses Global.

- *Hernández-Irizarry, R., Zendejas, B., Ali, S. M., & Farley, D. R. (2016). Optimizing training cost-effectiveness of simulation-based laparoscopic inguinal hernia repairs. *American Journal of Surgery*, *211*(2), 326-335. <https://doi.org/10.1016/j.amjsurg.2015.07.027>
- *Herrmann-Werner, A., Nikendei, C., Keifenheim, K., Bosse, H. M., Lund, F., Wagner, R., Celebi, N., Zipfel, S., & Weyrich, P. (2013). "Best practice" skills lab training vs. a "see one, do one" approach in undergraduate medical education: an RCT on students' long-term ability to perform procedural clinical skills. *PLoS ONE*, *8*(9), e76354. <https://doi.org/10.1371/journal.pone.0076354>
- Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, *327*(7414), 557-560. <https://doi.org/10.1136/bmj.327.7414.557>
- Higgins, M., Madan, C., & Patel, R. (2021). Development and decay of procedural skills in surgery: A systematic review of the effectiveness of simulation-based medical education interventions. *The Surgeon*, *19*(4), e67-e77. <https://doi.org/10.1016/j.surge.2020.07.013>
- Hill, L. B. (1934). A quarter century of delayed recall. *The Pedagogical Seminary and Journal of Genetic Psychology*, *44*(1), 231-238. <https://doi.org/10.1080/08856559.1934.10532492>
- Hill, L. B. (1957). A second quarter century of delayed recall or relearning at 80. *Journal of Educational Psychology*, *48*(2), 65-68. <https://doi.org/10.1037/h0043037>
- Hill, L. B., Rejall, A. E., & Thorndike, E. L. (1913). Practice in the case of typewriting. *The Journal of Genetic Psychology*, *20*, 516-529. <https://doi.org/10.1080/00221325.1991.9914706>
- *H'Mida, C., Kalyuga, S., Souissi, N., Rekik, G., Jarraya, M., & Khacharem, A. (2022). Is the human movement effect stable over time? The effects of presentation format on acquisition and retention of a motor skill. *Journal of Computer Assisted Learning*, *38*(1), 167-177. <https://doi.org/10.1111/jcal.12598>
- *Hodges, N. J., Edwards, C., Luttin, S., & Bowcock, A. (2011). Learning from the experts: Gaining insights into best practice during the acquisition of three novel motor skills. *Research Quarterly for Exercise and Sport*, *82*(2), 178-187. <https://doi.org/10.1080/02701367.2011.10599745>

- *Hodges, N. J., Hayes, S. J., Eaves, D. L., Horn, R. R., & Williams, A. M. (2006). End-point trajectory matching as a method for teaching kicking skills. *International Journal of Sport Psychology*, 37(2/3), 230-247.
- *Hoefer, S. H., Sterz, J., Bender, B., Stefanescu, M. C., Theis, M., Walcher, F., Sader, R., & Ruessler, M. (2017). Conveying practical clinical skills with the help of teaching associates-a randomised trial with focus on the long term learning retention. *BMC Medical Education*, 17(1), 65. <https://doi.org/10.1186/s12909-017-0892-5>
- *Holmgren, J. E., & Litton Industries. Litton-Mellonics Systems Development Division. (1979). *Training effectiveness and retention of training extension course (TEC) instruction in the combat arms* (Research Report 1208). Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences. <https://apps.dtic.mil/sti/citations/ADA069942>
- *Homan, C. S., Viccellio, P., Thode, H. C., Jr., & Fisher, W. (1994). Evaluation of an emergency-procedure teaching laboratory for the development of proficiency in tube thoracostomy. *Academic Emergency Medicine*, 1(4), 382-387. <https://doi.org/10.1111/j.1553-2712.1994.tb02649.x>
- *Hooyman, A., Wulf, G., & Lewthwaite, R. (2014). Impacts of autonomy-supportive versus controlling instructional language on motor learning. *Human Movement Science*, 36, 190-198. <https://doi.org/10.1016/j.humov.2014.04.005>
- *Horn, R. R., Williams, A. M., Scott, M. A., & Hodges, N. J. (2005). Visual search and coordination changes in response to video and point-light demonstrations without KR. *Journal of Motor Behavior*, 37(4), 265-274.
- *Howells, N. R., Auplish, S., Hand, G. C., Gill, H. S., Carr, A. J., & Rees, J. L. (2009). Retention of arthroscopic shoulder skills learned with use of a simulator: Demonstration of a learning curve and loss of performance level after a time delay. *Journal of Bone and Joint Surgery*, 91(5), 1207-1213. <https://doi.org/10.2106/jbjs.H.00509>
- *Huet, M., Jacobs, D. M., Camachon, C., Missenard, O., Gray, R., & Montagne, G. (2011). The education of attention as explanation of variability of practice effects: Learning the final approach phase in a flight simulator. *Journal of Experimental Psychology: Human Perception and Performance*, 37(6), 1841-1854. <https://doi.org/10.1037/a0024386>

- *Hufford, L. E., & Adams, J. A. (1961). *The contribution of part-task training to the relearning of a flight maneuver* (Technical Report 297-2). Port Washington, NY: US Naval Training Device Center. <https://apps.dtic.mil/sti/citations/AD0259505>
- *Huinink, L. H., Bouwsema, H., Plettenburg, D. H., van der Sluis, C. K., & Bongers, R. M. (2016). Learning to use a body-powered prosthesis: changes in functionality and kinematics. *Journal of Neuroengineering and Rehabilitation*, *13*(1), 90. <https://doi.org/10.1186/s12984-016-0197-7>
- Hunter, J. E. & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park C: Sage.
- Hurlock, R. E., & Montague, W. E. (1982). Skill retention and its implications for navy tasks: An analytical review (NPRDC Special Report No. 82-21). San Diego, CA: Navy Personnel Research and Development Center. <https://apps.dtic.mil/sti/citations/ADA114211>
- *Hussain, S. J., Darling, W. G., & Cole, K. J. (2016). Recent history of effector use modulates practice-dependent changes in corticospinal excitability but not motor learning. *Brain Stimulation*, *9*(4), 584-593. <https://doi.org/10.1016/j.brs.2016.03.019>
- *Hwang, G. Y. (2003). *An examination of the impact of introducing greater contextual interference during practice on learning to golf putt* (Publication No. 3156756). [Doctoral dissertation, Texas A&M University]. ProQuest Dissertations & Theses Global.
- *Iezzi, E., Suppa, A., Conte, A., Agostino, R., Nardella, A., & Berardelli, A. (2010). Theta-burst stimulation over primary motor cortex degrades early motor learning. *European Journal of Neuroscience*, *31*(3), 585-592. <https://doi.org/10.1111/j.1460-9568.2010.07090.x>
- *Isbye, D. L., Høiby, P., Rasmussen, M. B., Sommer, J., Lippert, F. K., Ringsted, C., & Rasmussen, L. S. (2008). Voice advisory manikin versus instructor facilitated training in cardiopulmonary resuscitation. *Resuscitation*, *79*(1), 73-81. <https://doi.org/10.1016/j.resuscitation.2008.06.012>
- *Iserbyt, P., Elen, J., & Behets, D. (2009). The effect of self-assessment in reciprocal learning with task cards on the quality of CPR. *Acta Anaesthesiologica Belgica*, *60*(4), 239-245.

- *Iserbyt, P., & Mols, L. (2014). Retention of CPR skills and the effect of instructor expertise one year following reciprocal learning. *Acta Anaesthesiol Belg*, 65(1), 23-29.
- *Ishikura, T. (2005). Average kr schedule in learning of timing: Influence of length for summary knowledge of results and task complexity. *Perceptual and Motor Skills*, 101(3), 911-924. <https://doi.org/10.2466/pms.101.3.911-924>
- *Ishikura, T. (2008). Reduced relative frequency of knowledge of results without visual feedback in learning a golf-putting task. *Perceptual and Motor Skills*, 106(1), 225-233. <https://doi.org/10.2466/PMS.106.1.225-233>
- *Iwatsuki, T., & Regis, C. J. (2021). Relatively easy criteria for success enhances motor learning by altering perceived competence. *Perceptual and Motor Skills*, 128(2), 900-911. <https://doi.org/10.1177/0031512520981237>
- *Jacquet, C. (2002). *A study of the effects of virtual reality on the retention of training* (ADA401460). [Master's thesis, University of Central Florida]. Defense Technical Information Center. <https://apps.dtic.mil/sti/citations/ADA401460>
- *Jahn, G., & Krems, J. F. (2013). Skill acquisition with text-entry interfaces: particularly older users benefit from minimized information-processing demands. *Journal of Applied Gerontology*, 32(5), 605-626. <https://doi.org/10.1177/0733464811433485>
- *Jang, K., Kim, S. H., Oh, J. Y., & Mun, J. Y. (2021). Effectiveness of self-re-learning using video recordings of advanced life support on nursing students' knowledge, self-efficacy, and skills performance. *BMC Nursing*, 20(1), 52. <https://doi.org/10.1186/s12912-021-00573-8>
- *Jang, T. C., Ryoo, H. W., Moon, S., Ahn, J. Y., Lee, D. E., Lee, W. K., ... & Kim, J. H. (2020). Long-term benefits of chest compression-only cardiopulmonary resuscitation training using real-time visual feedback manikins: a randomized simulation study. *Clinical and Experimental Emergency Medicine*, 7(3), 206-212. <https://doi.org/10.15441/2Fceem.20.022>
- Kelc, R., Vogrin, M., & Kelc, J. (2020). Cognitive training for the prevention of skill decay in temporarily non-performing orthopedic surgeons. *Acta Orthopaedica*, 91(5), 523-526. <https://doi.org/10.1080/17453674.2020.1771520>

- *Janse, J. A., Goedegebuure, R. S. A., Veersema, S., Broekmans, F. J. M., & Schreuder, H. W. R. (2013). Hysteroscopic sterilization using a virtual reality simulator: Assessment of learning curve. *Journal of Minimally Invasive Gynecology*, 20(6), 775-782. <https://doi.org/10.1016/j.jmig.2013.04.016>
- *Janse, J. A., Tolman, C. J., Veersema, S., Broekmans, F. J. M., & Schreuder, H. W. R. (2014). Hysteroscopy training and learning curve of 30° camera navigation on a new box trainer: The HYSTT. *Gynecological Surgery*, 11(2), 67-73. <https://doi.org/10.1007/s10397-014-0833-9>
- *Jaud, C., Salleron, J., Cisse, C., Angioi-Duprez, K., Berrod, J. P., & Conart, J. B. (2021). EyeSi Surgical Simulator: validation of a proficiency-based test for assessment of vitreoretinal surgical skills. *Acta Ophthalmologica*, 99(4), 390-396. <https://doi.org/10.1111/aos.14628>
- *Jenkins, J. L. (2008). *The effect of different training protocols in helicopter underwater escape training on egress performance* (publication No.233786716). [Master's thesis, Dalhousie University]. University of Ottawa: Library and Archives Canada. <https://library-archives.canada.ca/eng/services/services-libraries/theses/Pages/item.aspx?idNumber=233786716>
- *Jiang, G., Chen, H., Wang, S., Zhou, Q., Li, X., Chen, K., & Sui, X. (2011). Learning curves and long-term outcome of simulation-based thoracentesis training for medical students. *BMC Medical Education*, 11(1), 39. <https://doi.org/10.1186/1472-6920-11-39>
- *Johnson, C. E., Kimble, L. P., Gunby, S. S., & Davis, A. H. (2020). Using deliberate practice and simulation for psychomotor skill competency acquisition and retention: a mixed-methods study. *Nurse Educator*, 45(3), 150-154. <https://doi.org/10.1097/NNE.0000000000000713>
- *Jones, M. B. (1986). *Nonimposed overpractice and skill retention* (ARI Research Note 86-85). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. <https://apps.dtic.mil/sti/citations/ADA168572>
- *Jordan, T. C. (1971). *Visual and proprioceptive feedback variables as factors in the learning and retention of a motor skill* (Publication No. 7205934). [Doctoral dissertation, Stanford University]. ProQuest Dissertations & Theses Global.
- *Jowett, N., LeBlanc, V., Xeroulis, G., MacRae, H., & Dubrowski, A. (2007). Surgical skill acquisition with self-directed practice using computer-based video training.

American Journal of Surgery, 193(2), 237-242.
<https://doi.org/10.1016/j.amjsurg.2006.11.003>

- *Jujo, S., Lee-Jayaram, J. J., Sakka, B. I., Nakahira, A., Kataoka, A., Izumo, M., Kusunose, K., Athinartrattanapong, N., Oikawa, S., & Berg, B. W. (2021). Pre-clinical medical student cardiac point-of-care ultrasound curriculum based on the American Society of Echocardiography recommendations: a pilot and feasibility study. *Pilot and Feasibility Studies*, 7(1), 175. <https://doi.org/10.1186/s40814-021-00910-3>

- *Kaas, A. L., Van De Ven, V., Reithler, J., & Goebel, R. (2013). Tactile perceptual learning: Learning curves and transfer to the contralateral finger. *Experimental Brain Research*, 224(3), 477-488. <https://doi.org/10.1007/s00221-012-3329-8>

- *Kaczorowski, J., Levitt, C., Hammond, M., Outerbridge, E., Grad, R., Rothman, A., & Graves, L. (1998). Retention of neonatal resuscitation skills and knowledge: a randomized controlled trial. *Family Medicine*, 30(10), 705-711.

- *Kahol, K., Ashby, A., Smith, M., & Ferrara, J. J. (2010). Quantitative evaluation of retention of surgical skills learned in simulation. *Journal of Surgical Education*, 67(6), 421-426. <https://doi.org/10.1016/j.jsurg.2010.05.005>

- *Kapoor, R., Sandoval, M. A., Avendaño, L., Cruz, A. T., Soto, M. A., Camp, E. A., & Crouse, H. L. (2016). Regional scale-up of an Emergency Triage Assessment and Treatment (ETAT) training programme from a referral hospital to primary care health centres in Guatemala. *Emergency Medical Journal*, 33(9), 611-617. <https://doi.org/10.1136/emered-2015-205057>

- *Karlinsky, A., & Hodges, N. J. (2019). Manipulations to practice organization of golf putting skills through interleaved matched or mismatched practice with a partner. *Human Movement Science*, 66, 231-240. <https://doi.org/10.1016/j.humov.2019.05.001>

- *Keetch, K. M., & Lee, T. D. (2007). The effect of self-regulated and experimenter-imposed practice schedules on motor learning for tasks of varying difficulty. *Research Quarterly for Exercise and Sport*, 78(5), 476-486. <https://doi.org/10.1080/02701367.2007.10599447>

- *Keh, N. C. (1992). *Students' use of teacher feedback during badminton instruction* (Publication No. 9302906). [Doctoral dissertation, Louisiana State University and Agricultural & Mechanical College]. ProQuest Dissertations & Theses Global.

- *Kernodle, M. W., Johnson, R., & Arnold, D. R. (2001). Verbal instruction for correcting errors versus such instructions plus videotape replay on learning the overhand throw. *Perceptual and Motor Skills*, 92(3 PART 2), 1039-1051. <https://doi.org/10.2466/pms.2001.92.3c.1039>
- *Khan, A., Singh, J., Neary, J. P., & Mang, C. S. (2022). Exercise effects on motor skill consolidation and intermuscular coherence depend on practice schedule. *Brain Sciences*, 12(4), 436. <https://doi.org/10.3390/brainsci12040436>
- *Khan, M. W., Lin, D., Marlow, N., Aintree, M., Babidge, W., Field, J., Hewett, P., & Maddern, G. (2014). Laparoscopic skills maintenance: A randomized trial of virtual reality and box trainer simulators. *Journal of Surgical Education*, 71(1), 79-84. <https://doi.org/10.1016/j.jsurg.2013.05.009>
- *Khanjari, Y., Arabameri, E., Shahbazi, M., Boroujeni, S. T., & Bahrami, F. (2020). Synergy patterns of brain activity during learning of the dart throwing skill with the dominant and non-dominant hand. *South African Journal for Research in Sport, Physical Education and Recreation*, 42(2), 49-66.
- *Kleinrensink, V. E. E., Lange, J. F., & Kleinrensink, G. J. (2021). Perfectionistic concerns are detrimental to skill learning for minimally invasive surgery. *Journal of Surgical Education*, 78(2), 630-637. <https://doi.org/10.1016/j.jsurg.2020.08.022>
- *Klostermann, A., & Hossner, E. J. (2018). The quiet eye and motor expertise: Explaining the "efficiency paradox". *Frontiers in Psychology*, 9, 104. <https://doi.org/10.3389/fpsyg.2018.00104>
- *Klostermann, A., Reinbold, F., & Kredel, R. (2022). Learning different task spaces: how explored density aligns the Quiet Eye. *Cognitive Processing*, 23(3), 449-458. <https://doi.org/10.1007/s10339-022-01090-5>
- *Knerr, C., Harris, J., O'Brien, D., Sticha, P., & Goldberg, S. L. (1984). *Armor procedural skills: Learning and retention*. apps.dtic.mil. <https://apps.dtic.mil/sti/citations/ADA153227>
- *Kniha, K., Bock, A., Peters, F., Heitzer, M., Hölzle, F., Raith, S., Modabber, A., & Möhlhenrich, S. C. (2022). Guided discovery learning: A follow-up study of try-it-yourself surgery and subsequent video-assisted teaching for oral surgical skills training. *European Journal of Dental Education*. <https://doi.org/10.1111/eje.12772>

- *Kim, J., Singer, R. N., & Radlo, S. J. (1996). Degree of cognitive demands in psychomotor tasks and the effects of the five-step strategy on achievement. *Human Performance*, 9(2), 155-169. https://doi.org/10.1207/s15327043hup0902_4
- *Kim, J. W., Koubek, R. J., & Ritter, F. E. (2007). Investigation of procedural skills degradation from different modalities. In *Proceedings of the 8th international conference on cognitive modeling* (pp. 255-260). Ann Arbor, MI: Psychology Press.
- *Kim, T. (2020). *Applying anodal transcranial direct current stimulation at BA6 during repetitive practice enhances motor learning by improving encoding and post-practice consolidation*. [Doctoral dissertation, Texas A&M University]. Texas A&M University Libraries. <https://hdl.handle.net/1969.1/191713>
- *Kim, T., Chen, J., Verwey, W. B., & Wright, D. L. (2018). Improving novel motor learning through prior high contextual interference training. *Acta Psychologica*, 182, 55-64. <https://doi.org/10.1016/j.actpsy.2017.11.005>
- *Kim, T., Frank, C., & Schack, T. (2017). A systematic investigation of the effect of action observation training and motor imagery training on the development of mental representation structure and skill performance. *Frontiers in Human Neuroscience*, 11, 499. <https://doi.org/10.3389/fnhum.2017.00499>
- *Kingsley, J. L. (1970). *The effectiveness of teaching large motor skills by the spaced-unit method as compared with the continuous-unit method* (Publication No. 7023285). [Doctoral dissertation, New York University]. ProQuest Dissertations & Theses Global.
- *Knerr, C. M., Harris, J. H., O'Brien, B. K., Sticha, P. J., & Goldberg, S. L. (1984). *Armor procedural skills: Learning and retention* (Technical Report 621). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. <https://apps.dtic.mil/sti/citations/ADA153227>
- *Kolozsvari, N. O., Kaneva, P., Brace, C., Chartrand, G., Vaillancourt, M., Cao, J., Banaszek, D., Demyttenaere, S., Vassiliou, M. C., Fried, G. M., & Feldman, L. S. (2011). Mastery versus the standard proficiency target for basic laparoscopic skill training: effect on skill transfer and retention. *Surgical Endoscopy*, 25(7), 2063-2070. <https://doi.org/10.1007/s00464-011-1743-9>

- *Kopacek, K. B., Dopp, A. L., Dopp, J. M., Vardeny, O., & Sims, J. J. (2010). Pharmacy students' retention of knowledge and skills following training in automated external defibrillator use. *American Journal of Pharmaceutical Education*, 74(6), 109. <https://doi.org/10.5688/aj7406109>
- *Korman, M., Doyon, J., Doljansky, J., Carrier, J., Dagan, Y., & Karni, A. (2007). Daytime sleep condenses the time course of motor memory consolidation. *Nature Neuroscience*, 10(9), 1206-1213. <https://doi.org/10.1038/nn1959>
- Kovacs, G., Bullock, G., Ackroyd-Stolarz, S., Cain, E., & Petrie, D. (2000). A randomized controlled trial on the effect of educational interventions in promoting airway management skill maintenance. *Annals of Emergency Medicine*, 36(4), 301-309. <https://doi.org/10.1067/mem.2000.109339>
- *Kraeutner, S. N., Stratas, A., McArthur, J. L., Helmick, C. A., Westwood, D. A., & Boe, S. G. (2020). Neural and behavioral outcomes differ following equivalent bouts of motor imagery or physical practice. *Journal of Cognitive Neuroscience*, 32(8), 1590-1606. https://doi.org/10.1162/jocn_a_01575
- *Kratzig, G. P. (2016). *Skill retention: A test of the effects of overlearning and skill retention interval on maintenance of infrequently used complex skills* (Publication No. 28141744). [Doctoral dissertation, The University of Regina]. ProQuest Dissertations & Theses Global.
- *Kruglikova, I., Grantcharov, T. P., Drewes, A. M., & Funch-Jensen, P. (2010). The impact of constructive feedback on training in gastrointestinal endoscopy using high-fidelity Virtual-Reality simulation: a randomised controlled trial. *Gut*, 59(2), 181-185. <https://doi.org/10.1136/gut.2009.191825>
- *Kudreviciene, A., Nadisauskiene, R. J., Tameliene, R., Tamelis, A., Nedzelskiene, I., Dobožinskas, P., & Vaitkaitis, D. (2019). Initial neonatal resuscitation: skill retention after the implementation of the novel 24/7 HybridLab® learning system. *Journal of Maternal-Fetal and Neonatal Medicine*, 32(8), 1230-1237. <https://doi.org/10.1080/14767058.2017.1402881>
- *Kunaratnam, N. (2018). *Effect of anodal transcranial direct current stimulation on motor learning in healthy individuals* (Publication No. 10936163). [Master's thesis, University of Toronto]. ProQuest Dissertations & Theses Global.
- *Kyle Leming, J., Dorman, K., Brydges, R., Carnahan, H., & Dubrowski, A. (2007). Tensiometry as a measure of improvement in knot quality in undergraduate

medical students. *Advances in Health Sciences Education*, 12, 331-344.
<https://doi.org/10.1007/s10459-006-9005-7>

*Laack, T. A., Dong, Y., Goyal, D. G., Sadosty, A. T., Suri, H. S., & Dunn, W. F. (2014). Short-term and long-term impact of the central line workshop on resident clinical performance during simulated central line placement. *Simulation in Healthcare*, 9(4), 228-233. <https://doi.org/10.1097/SIH.0000000000000015>

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 863.
<https://doi.org/10.3389/fpsyg.2013.00863>

*Lampton, D., Bliss, J., & Meert, M. (1992). *The Effects of Differences between Practice and Test Criteria on Transfer and Retention of a Simulated Tank Gunnery Task* (ARI Technical Report 949). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
<https://apps.dtic.mil/sti/citations/ADA251867>

*Land, W. M., Frank, C., & Schack, T. (2014). The influence of attentional focus on the development of skill representation in a complex action. *Psychology of Sport and Exercise*, 15(1), 30-38. <https://doi.org/10.1016/j.psychsport.2013.09.006>

*Larrivée, S., Rodger, R., Larouche, P., Leiter, J., Jelic, T., & MacDonald, P. (2019). Orthopaedic residents demonstrate retention of point of care ultrasound knowledge after a brief educational session: a quasi experimental study. *BMC Medical Education*, 19(1), 474. <https://doi.org/10.1186/s12909-019-1916-0>

*Latif, R. K., Bautista, A., Duan, X., Neamtu, A., Wu, D., Wadhwa, A., & Akça, O. (2016). Teaching basic fiberoptic intubation skills in a simulator: initial learning and skills decay. *Journal of Anesthesia*, 30(1), 12-19.
<https://doi.org/10.1007/s00540-015-2091-z>

*Latif, R. K., VanHorne, E. M., Kandadai, S. K., Bautista, A. F., Neamtu, A., Wadhwa, A., Carter, M. B., Ziegler, C. H., Memon, M. F., & Akça, O. (2016). Teaching basic lung isolation skills on human anatomy simulator: attainment and retention of lung isolation skills. *BMC Anesthesiology*, 16, 7.
<https://doi.org/10.1186/s12871-015-0169-7>

Lau, J., Ioannidis, J. P. A., Terrin, N., Schmid, C. H., & Olkin, I. (2006). The case of the misleading funnel plot. *British Medical Journal*, 333(7568), 597-600.
<https://doi.org/10.1136/bmj.333.7568.597>

- *Lazarus, J. A. C., & Haynes, J. (1997). Isometric pinch force control and learning in older adults. *Experimental Aging Research*, 23(2), 179-199.
<https://doi.org/10.1080/03610739708254032>
- *Lean, L. L., Hong, R. Y. S., & Ti, L. K. (2017). End-task versus in-task feedback to increase procedural learning retention during spinal anaesthesia training of novices. *Advances in Health Sciences Education*, 22(3), 713-721.
<https://doi.org/10.1007/s10459-016-9703-8>
- *LeBel, M. E., Haverstock, J., Cristancho, S., van Eimeren, L., & Buckingham, G. (2018). Observational learning during simulation-based training in arthroscopy: Is it useful to novices? *Journal of Surgical Education*, 75(1), 222-230.
<https://doi.org/10.1016/j.jsurg.2017.06.005>
- *Lee, H., & Choi, S. (2014, February). Combining haptic guidance and haptic disturbance: an initial study of hybrid haptic assistance for virtual steering task. In *2014 IEEE haptics symposium (HAPTICS)* (pp. 159-165). IEEE.
<https://doi.org/10.1109/HAPTICS.2014.6775449>
- *Lee, M. M. H., Chan, C. N., Lau, B. Y. T., & Ma, T. W. L. (2017). Randomised controlled study to assess skill retention at 6 vs 12 months after simulation training in shoulder dystocia. *BMJ Simulation & Technology Enhanced Learning*, 3(4), 142-148. <https://doi.org/10.1136/bmjstel-2017-000195>
- *Lee, T. D., Wishart, L. R., Cunningham, S., & Carnahan, H. (1997). Modeled timing information during random practice eliminates the contextual interference effect. *Research Quarterly for Exercise and Sport*, 68(1), 100-105.
<https://doi.org/10.1080/02701367.1997.10608871>
- *Leinen, P., Panzer, S., & Shea, C. H. (2016). Hemispheric asymmetries of a motor memory in a recognition test after learning a movement sequence. *Acta Psychologica*, 171, 36-46. <https://doi.org/10.1016/j.actpsy.2016.09.007>
- *Lelievre, N., St. Germain, L., & Ste-Marie, D. M. (2021). Varied speeds of video demonstration do not influence the learning of a dance skill. *Human Movement Science*, 75, 102749. <https://doi.org/10.1016/j.humov.2020.102749>
- *Lengetti, E., Kronk, R., Ulmer, K. W., Wilf, K., Murphy, D., Rosanelli, M., & Taylor, A. (2018). An innovative approach to educating nurses to clinical competence: A randomized controlled trial. *Nurse Education in Practice*, 33, 159-163.
<https://doi.org/10.1016/j.nepr.2018.08.007>

- *Lewthwaite, R., Chiviawosky, S., Drews, R., & Wulf, G. (2015). Choose to move: The motivational impact of autonomy support on motor learning. *Psychonomic Bulletin and Review*, 22(5), 1383-1388. <https://doi.org/10.3758/s13423-015-0814-7>
- *Li, G., Li, H., Pu, J., Wan, F., & Hu, Y. (2021). Effect of brain alpha oscillation on the performance in laparoscopic skills simulator training. *Surgical Endoscopy*, 35(2), 584-592. <https://doi.org/10.1007/s00464-020-07419-5>
- *Li, Q., Zhou, R. H., Liu, J., Lin, J., Ma, E. L., Liang, P., Shi, T. W., Fang, L. Q., & Xiao, H. (2013). Pre-training evaluation and feedback improved skills retention of basic life support in medical students. *Resuscitation*, 84(9), 1274-1278. <https://doi.org/10.1016/j.resuscitation.2013.04.017>
- *Li, W., Meng, X., Zhang, K. J., Yang, Z., Feng, Z., Tong, K., & Tian, J. (2022). Meditation using a mobile app improves surgery trainee performance: A simulation-based randomized controlled trial. *Arthroscopy*. <https://doi.org/10.1016/j.arthro.2022.09.008>
- *Li, W., Zhang, K. J., Yao, S., Xie, X., Han, W., Xiong, W. B., & Tian, J. (2020). Simulation-based arthroscopic skills using a spaced retraining schedule reduces short-term task completion time and camera path length. *Arthroscopy*, 36(11), 2866-2872. <https://doi.org/10.1016/j.arthro.2020.05.040>
- *Lin, L., Ni, S., Liu, Y., Xue, J., Ma, B., Xiong, D., Zhao, Y., & Jin, X. (2022). Effect of peer videorecording feedback CPR training on students' practical CPR skills: a randomized controlled manikin study. *BMC Medical Education*, 22(1), 484. <https://doi.org/10.1186/s12909-022-03563-9>
- *Lin-Martore, M., Olvera, M. P., Kornblith, A. E., Zapala, M., Addo, N., Lin, M., & Werner, H. C. (2021). Evaluating a web-based point-of-care ultrasound curriculum for the diagnosis of intussusception. *AEM Educational Training*, 5(3), e10526. <https://doi.org/10.1002/aet2.10526>
- Linde, A. S., Caridha, J., & Kunkler, K. J. (2018). Skills decay in military medical training: a meta-synthesis of research outcomes. *Military Medicine*, 183(1-2), e40-e44. <https://doi.org/10.1093/milmed/usx023>
- Linde, A. S., & Miller, G. T. (2019). Applications of future technologies to detect skill decay and improve procedural performance. *Military Medicine*, 184(Supplement_1), 72-77. <https://doi.org/10.1093/milmed/usy385>

- *von Lindern, A. D., & Fairbrother, J. T. (2022). Reduction of feedback availability limits self-control effects. *Frontiers in Sports and Active Living*, 4, 816571. <https://doi.org/10.3389/fspor.2022.816571>
- *Linsk, A. M., Monden, K. R., Sankaranarayanan, G., Ahn, W., Jones, D. B., De, S., Schwaitzberg, S. D., & Cao, C. G. L. (2018). Validation of the VBLaST pattern cutting task: a learning curve study. *Surgical Endoscopy*, 32(4), 1990-2002. <https://doi.org/10.1007/s00464-017-5895-0>
- Lionetti, P. (2012). *Transfer of training: 1988-2077 with the practitioner in mind* (Publication No. 3547780). [Doctoral dissertation, Pepperdine University]. ProQuest Dissertations & Theses Global.
- *Liu, Y. T., & Newell, K. M. (2015). S-Shaped motor learning and nonequilibrium phase transitions. *Journal of Experimental Psychology: Human Perception and Performance*, 41(2), 403-414. <https://doi.org/10.1037/a0038812>
- *Liu, Z. J., Yi, J., Chen, W. Y., Zhang, X. H., & Huang, Y. G. (2017). Comparison of learning performance of 2 intubating laryngeal mask airways in novice: A randomized crossover manikin study. *Medicine*, 96(19), e6905. <https://doi.org/10.1097/md.0000000000006905>
- Lohre, R., Warner, J. J., Morrey, B. R., Athwal, G. S., Morrey, M. E., Mazzocca, A. D., & Goel, D. P. (2021). Mitigating surgical skill decay in orthopaedics using virtual simulation learning. *JAAOS Global Research & Reviews*, 5(10), e21. <https://doi.org/10.5435/JAAOSGlobal-D-21-00193>
- *Luckey-Smith, K., High, K., & Cole, E. (2020). Effectiveness of surgical airway training laboratory and assessment of skill and knowledge fade in surgical airway establishment among prehospital providers. *Air Medical Journal*, 39(5), 369-373. <https://doi.org/10.1016/j.amj.2020.05.019>
- *Luken, M., & Yancosek, K. (2017). Effects of an occupational therapy hand dominance transfer intervention for soldiers with crossed hand-eye dominance. *Journal of Motor Behavior*, 49(1), 78-87. <https://doi.org/10.1080/00222895.2016.1191420>
- Lyon, R. M. (1977). *Identification of classic studies in motor learning* (Publication No DP29724). [Doctoral dissertation, University of Southern California]. ProQuest Dissertations & Theses Global.

- *Maagaard, M., Sorensen, J. L., Oestergaard, J., Dalsgaard, T., Grantcharov, T. P., Ottesen, B. S., & Larsen, C. R. (2011). Retention of laparoscopic procedural skills acquired on a virtual-reality surgical trainer. *Surgical Endoscopy*, *25*(3), 722-727. <https://doi.org/10.1007/s00464-010-1233-5>
- *Mackenzie, C. F., Shackelford, S. A., Tisherman, S. A., Yang, S., Puche, A., Elster, E. A., & Bowyer, M. W. (2019). Critical errors in infrequently performed trauma procedures after training. *Surgery*, *166*(5), 835-843. <https://doi.org/10.1016/j.surg.2019.05.031>
- *Macluskey, M., & Hanson, C. (2011). The retention of suturing skills in dental undergraduates. *European Journal of Dental Education*, *15*(1), 42-46. <https://doi.org/10.1111/j.1600-0579.2010.00632.x>
- *Madden, C. (2006). Undergraduate nursing students' acquisition and retention of CPR knowledge and skills. *Nurse Education Today*, *26*(3), 218-227. <https://doi.org/10.1016/j.nedt.2005.10.003>
- *Maeda, R. S., McGee, S. E., & Marigold, D. S. (2018). Long-term retention and reconsolidation of a visuomotor memory. *Neurobiology of Learning and Memory*, *155*, 313-321. <https://doi.org/10.1016/j.nlm.2018.08.020>
- *Maertens, H., Aggarwal, R., Moreels, N., Vermassen, F., & Van Herzelee, I. (2017). A proficiency based stepwise endovascular curricular training (PROSPECT) program enhances operative performance in real life: A randomised controlled trial. *European Journal of Vascular and Endovascular Surgery*, *54*(3), 387-396. <https://doi.org/10.1016/j.ejvs.2017.06.011>
- *Manzone, J. C. (2015). *Rethinking transfer: How supervision and support during self-regulated learning impacts medical trainees' preparation for future learning in a simulation setting* Publication No. 1604640). [Master's thesis, University of Toronto]. ProQuest Dissertations & Theses Global.
- Mairaj, A., Baba, A. I., & Javaid, A. Y. (2019). Application specific drone simulators: Recent advances and challenges. *Simulation Modelling Practice and Theory*, *94*, 100-117. <https://doi.org/10.1016/j.simpat.2019.01.004>
- *Marchal-Crespo, L., McHughen, S., Cramer, S. C., & Reinkensmeyer, D. J. (2010). The effect of haptic guidance, aging, and initial skill level on motor learning of a steering task. *Experimental Brain Research*, *201*(2), 209-220. <https://doi.org/10.1007/s00221-009-2026-8>

- *Marchand, A. A., Mendoza, L., Dugas, C., Descarreaux, M., & Pagé, I. (2017). Effects of practice variability on spinal manipulation learning. *Journal Chiropractic Education, 31*(2), 90-95. <https://doi.org/10.7899/jce-16-8>
- *Marmie, W. R., & Healy, A. F. (1995). The long-term retention of a complex skill. In A. F. Healy & L. E. Bourne, Jr. (Eds.), *Learning and memory of knowledge and skills: Durability and specificity*. (pp. 30-65). Sage Publications, Inc. <https://doi.org/10.4135/9781483326887.n2>
- *Marques, P. G., & Corrêa, U. C. (2016). The effect of learner's control of self-observation strategies on learning of front crawl. *Acta Psychologica, 164*, 151-156. <https://doi.org/10.1016/j.actpsy.2016.01.006>
- *Marschalek, J., Kuessel, L., Stammeler-Safar, M., Kiss, H., Ott, J., & Husslein, H. (2022). Comparison of a practice-based versus theory-based training program for conducting vacuum-assisted deliveries: a randomized-controlled trial. *Archives of Gynecology and Obstetrics, 305*(2), 365-372. <https://doi.org/10.1007/s00404-021-06159-8>
- *Massing, M., Blandin, Y., & Panzer, S. (2016). Magnifying visual target information and the role of eye movements in motor sequence learning. *Acta Psychologica, 163*, 59-64. <https://doi.org/10.1016/j.actpsy.2015.11.004>
- Mathur, M. B., & VanderWeele, T. J. (2021). Estimating publication bias in meta-analyses of peer-reviewed studies: A meta-meta-analysis across disciplines and journal tiers. *Research Synthesis Methods, 12*(2), 176-191. <https://doi.org/10.1002/jrsm.1464>
- *McAlister, R. (2006). *The effects of attentional focus instructions on simulated upper extremity amputees' movement kinematics when learning a novel functional task* [Doctoral dissertation, Auburn University]. Auburn University Repository. <https://etd.auburn.edu/handle/10415/588>
- McBride, E. R., & Rothstein, A. L. (1979). Mental and physical practice and the learning and retention of open and closed skills. *Perceptual and Motor Skills, 49*(2), 359-365. <https://doi.org/10.2466/pms.1979.49.2.359>
- *McCann, P., Lavalley, D., & Lavalley, R. M. (2001). The effect of pre-shot routines on golf wedge shot performance. *European Journal of Sport Science, 1*(5), 1-10. <https://doi.org/10.1080/17461390100071503>

- *McCullagh, P., & Little, W. S. (1990). Demonstrations and knowledge of results in motor skill acquisition. *Perceptual and Motor Skills*, 71(3 Pt 1), 735-742.
<https://doi.org/10.2466/pms.1990.71.3.735>
- *McKay, B., & Ste-Marie, D. M. (2022). Autonomy support via instructionally irrelevant choice not beneficial for motor performance or learning. *Research Quarterly for Exercise and Sport*, 93(1), 64-76.
<https://doi.org/10.1080/02701367.2020.1795056>
- *McRae, M., Patterson, J. T., & Hansen, S. (2015). Examining the preferred self-controlled KR schedules of learners and peers during motor skill learning. *Journal of Motor Behavior*, 47(6), 527-534.
<https://doi.org/10.1080/00222895.2015.1020357>
- *Mummert, D. (2006). Long-term effects of type of practice on the learning and transfer of a complex motor skill. *Perceptual and Motor Skills*, 103(3), 912-916.
<https://doi.org/10.2466/PMS.103.7.912-916>
- *Mengelkoch, R. F., Adams, J. A., & Gainer, C. A. (1971). The forgetting of instrument flying skills. *Human Factors*, 13(5), 397-405.
<https://doi.org/10.1177/001872087101300502>
- Meyers, J. L. (1967). Retention of balance coordination learning as influenced by extended lay-offs. *Research Quarterly. American Association for Health, Physical Education and Recreation*, 38(1), 72-78.
<https://doi.org/10.1080/10671188.1967.10614805>
- *Mikhaeil-Demo, Y., Barsuk, J. H., Culler, G. W., Bega, D., Salzman, D. H., Cohen, E. R., Templer, J. W., & Gerard, E. E. (2020). Use of a simulation-based mastery learning curriculum for neurology residents to improve the identification and management of status epilepticus. *Epilepsy & Behavior*, 111.
<https://doi.org/10.1016/j.yebeh.2020.107247>
- Moazed, F., Cohen, E. R., Furiasse, N., Singer, B., Corbridge, T. C., McGaghie, W. C., & Wayne, D. B. (2013). Retention of critical care skills after simulation-based mastery learning. *Journal of Graduate Medical Education*, 5(3), 458-463.
<https://doi.org/10.4300/JGME-D-13-00033.1>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA

statement. *Annals of Internal Medicine*, 151(4), 264-269.
<https://doi.org/10.7326/0003-4819-151-4-200908180-00135>

- *Molinas, C. R., & Campo, R. (2016). Retention of laparoscopic psychomotor skills after a structured training program depends on the quality of the training and on the complexity of the task. *Gynecological Surgery*, 13(4), 395-402.
<https://doi.org/10.1007/s10397-016-0962-4>

- *Mooney, R. A., Cirillo, J., Stinear, C. M., & Byblow, W. D. (2020). Neurophysiology of motor skill learning in chronic stroke. *Clinical Neurophysiology*, 131(4), 791-798.
<https://doi.org/10.1016/j.clinph.2019.12.410>

- *Moore, L. J., Wilson, M. R., Waine, E., Masters, R. S. W., McGrath, J. S., & Vine, S. J. (2015). Robotic technology results in faster and more robust surgical skill acquisition than traditional laparoscopy. *Journal of Robotic Surgery*, 9(1), 67-73.
<https://doi.org/10.1007/s11701-014-0493-9>

- *Moradi, J. (2020). Benefits of a guided motor-mental preperformance routine on learning the basketball free throw. *Perceptual and Motor Skills*, 127(1), 248-262.
<https://doi.org/10.1177/0031512519870648>

- *Muñoz, G. J., Cortéz, D. A., Álvarez, C. B., Raggio, J. A., Concha, A., Rojas, F. I., Arthur, W., Jr., Fischer, B. M., & Rodriguez, S. (2022). After-action reviews and long-term performance: An experimental examination in the context of an emergency simulation. *Human Factors*, 64(4), 760-778.
<https://doi.org/10.1177/0018720820958848>

- *Munzert, J., Maurer, H., & Reiser, M. (2014). Verbal-motor attention-focusing instructions influence kinematics and performance on a golf-putting task. *Journal of Motor Behavior*, 46(5), 309-318.
<https://doi.org/10.1080/00222895.2014.912197>

- *Murcia-López, M., & Steed, A. (2018). A comparison of virtual and physical training transfer of bimanual assembly tasks. *IEEE Transactions on Visualization and Computer Graphics*, 24(4), 1574-1583.
<https://doi.org/10.1109/TVCG.2018.2793638>

- *Naimo, M. A., Zourdos, M. C., Wilson, J. M., Kim, J. S., Ward, E. G., Eccles, D. W., & Panton, L. B. (2013). Contextual interference effects on the acquisition of skill and strength of the bench press. *Human Movement Science*, 32(3), 472-484.
<https://doi.org/10.1016/j.humov.2013.02.002>

Naylor, J. C., & Briggs, G. E. (1961). *Long-term retention of learned skills: A review of the literature* (Tech. Rep. No. 61-390). Wright-Patterson Air Force Base, OH: Aerospace Medical Laboratory, Aeronautical Systems Division.
<https://pubmed.ncbi.nlm.nih.gov/24546761/>

National Center for O*NET Development. *O*NET OnLine*. Retrieved July 11, 2023, from <https://www.onetonline.org/>

*Neville, K. M., & Trempe, M. (2017). Serial practice impairs motor skill consolidation. *Experimental Brain Research*, 235(9), 2601-2613.
<https://doi.org/10.1007/s00221-017-4992-6>

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Hillsdale, NJ: Erlbaum.

Niebuhr, D. W., Page, W. F., Cowan, D. N., Urban, N., Gubata, M. E., & Richard, P. (2013). Cost-effectiveness analysis of the US Army Assessment of Recruit Motivation and Strength (ARMS) program. *Military Medicine*, 178(10), 1102-1110. <https://doi.org/10.7205/MILMED-D-13-00108>

*Nimbalkar, A., Patel, D., Kungwani, A., Phatak, A., Vasa, R., & Nimbalkar, S. (2015). Randomized control trial of high fidelity vs low fidelity simulation for training undergraduate students in neonatal resuscitation. *BMC Research Notes*, 8, 636.
<https://doi.org/10.1186/s13104-015-1623-9>

*Nippita, S., Haviland, M. J., Voit, S. F., Perez-Peralta, J., Hacker, M. R., & Paul, M. E. (2018). Randomized trial of high- and low-fidelity simulation to teach intrauterine contraception placement. *American Journal of Obstetrics and Gynecology*, 218(2), 258.e251-258.e211. <https://doi.org/10.1016/j.ajog.2017.11.553>

*Nitsche, J. F., Knupp, R. J., Tooze, J. A., Sun, C. H., Jin, S., & Brost, B. C. (2019). Using simulation to define the learning curve for the digital cervical examination. *Obstetrics & Gynecology*, 134 Suppl 1, 1s-8s.
<https://doi.org/10.1097/aog.0000000000003433>

*Nourkami-Tutdibi, N., Hilleke, A. B., Zemlin, M., Wagenpfeil, G., & Tutdibi, E. (2020). Novel modified Peyton's approach for knowledge retention on newborn life support training in medical students. *Acta Paediatrica, International Journal of Paediatrics*, 109(8), 1570-1579. <https://doi.org/10.1111/apa.15198>

- *Nourouzpour, N., Salomonczyk, D., Cressman, E. K., & Henriques, D. Y. P. (2015). Retention of proprioceptive recalibration following visuomotor adaptation. *Experimental Brain Research*, 233(3), 1019-1029. <https://doi.org/10.1007/s00221-014-4176-6>
- *Nourrit, D., Delignières, D., Caillou, N., Deschamps, T., & Lauriot, B. (2003). On discontinuities in motor learning: A longitudinal study of complex skill acquisition on a ski-simulator. *Journal of Motor Behavior*, 35(2), 151-170. <https://doi.org/10.1080/00222890309602130>
- *Nourrit-Lucas, D., Zelic, G., Deschamps, T., Hilpron, M., & Delignières, D. (2013). Persistent coordination patterns in a complex task after 10 years delay. Subtitle: How validate the old saying "Once you have learned how to ride a bicycle, you never forget!". *Human Movement Science*, 32(6), 1365-1378. <https://doi.org/10.1016/j.humov.2013.07.005>
- *Nousiainen, M. T., Omoto, D. M., Zingg, P. O., Weil, Y. A., Mardam-Bey, S. W., & Eward, W. C. (2013). Training femoral neck screw insertion skills to surgical trainees: computer-assisted surgery versus conventional fluoroscopic technique. *Journal of Orthopaedic Trauma*, 27(2), 87-92. <https://doi.org/10.1097/BOT.0b013e3182604b49>
- *Oermann, M. H., Kardong-Edgren, S. E., Odom-Maryon, T., & Roberts, C. J. (2014). Effects of practice on competency in single-rescuer cardiopulmonary resuscitation. *Medsurg Nursing*, 23(1), 22-28.
- *O'Donnell, A. M., Dansereau, D. F., Hall, R. H., Skaggs, L. P., Hythecker, V. I., Peel, J. L., & Rewey, K. L. (1990). Learning concrete procedures: Effects of processing strategies and cooperative learning. *Journal of Educational Psychology*, 82(1), 171-177. <https://doi.org/10.1037/0022-0663.82.1.171>
- Ong, A. Y., Naughton, A., Stone, N., Aslam, S., & Higham, A. (2021). Implementing a mental practice training protocol to prevent decay of surgical skills during COVID-19 and beyond. *Journal of Cataract & Refractive Surgery*, 47(12), 1596-1597. <https://doi.org/10.1097/j.jcrs.0000000000000647>
- *Ong, N. T., Lohse, K. R., & Hodges, N. J. (2015). Manipulating target size influences perceptions of success when learning a dart-throwing skill but does not impact retention. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.01378>

- *van Ooteghem, K., Frank, J. S., Allard, F., & Horak, F. B. (2010). Aging does not affect generalized postural motor learning in response to variable amplitude oscillations of the support surface. *Experimental Brain Research*, *204*, 505-514. <https://doi.org/10.1007/s00221-010-2316-1>
- *Orlando, M. S., Thomaier, L., Abernethy, M. G., & Chen, C. C. G. (2017). Retention of laparoscopic and robotic skills among medical students: a randomized controlled trial. *Surgical Endoscopy*, *31*(8), 3306-3312. <https://doi.org/10.1007/s00464-016-5363-2>
- Orledge, J., Phillips, W. J., Murray, W. B., & Lerant, A. (2012). The use of simulation in healthcare: from systems issues, to team building, to task training, to education and high stakes examinations. *Current Opinion in Critical Care*, *18*(4), 326-332. <https://doi.org/10.1097/MCC.0b013e328353fb49>
- *Palmer, K., Chiviacowsky, S., & Wulf, G. (2016). Enhanced expectancies facilitate golf putting. *Psychology of Sport and Exercise*, *22*, 229-232. <https://doi.org/10.1016/j.psychsport.2015.08.009>
- *Panchuk, D., Spittle, M., Johnston, N., & Spittle, S. (2013). Effect of practice distribution and experience on the performance and retention of a discrete sport skill. *Perceptual and Motor Skills*, *116*(3), 750-760. <https://doi.org/10.2466/23.30.PMS.116.3.750-760>
- *Panzer, S., Gruetzmacher, N., Ellenbürger, T., & Shea, C. H. (2014). Interlimb practice and aging: Coding a simple movement sequence. *Experimental Aging Research*, *40*(1), 107-128. <https://doi.org/10.1080/0361073X.2014.857566>
- *Panzer, S., Gruetzmacher, N., Fries, U., Krueger, M., & Shea, C. H. (2011). Age-related effects in interlimb practice on coding complex movement sequences. *Human Movement Science*, *30*(3), 459-474. <https://doi.org/10.1016/j.humov.2010.11.003>
- *Panzer, S., Krueger, M., Muehlbauer, T., Kovacs, A. J., & Shea, C. H. (2009). Inter-manual transfer and practice: Coding of simple motor sequences. *Acta Psychologica*, *131*(2), 99-109. <https://doi.org/10.1016/j.actpsy.2009.03.004>
- *Pasquier, M., Cheron, C., Dugas, C., Lardon, A., & Descarreaux, M. (2017). The effect of augmented feedback and expertise on spinal manipulation skills: An experimental study. *Journal of Manipulative and Physiological Therapeutics*, *40*(6), 404-410. <https://doi.org/10.1016/j.jmpt.2017.03.010>

- *Partiprajak, S., & Thongpo, P. (2016). Retention of basic life support knowledge, self-efficacy and chest compression performance in Thai undergraduate nursing students. *Nurse Education in Practice*, 16(1), 235-241. <https://doi.org/10.1016/j.nepr.2015.08.012>
- *Patocka, C., Cheng, A., Sibbald, M., Duff, J. P., Lai, A., Lee-Nobbee, P., Levin, H., Varshney, T., Weber, B., & Bhanji, F. (2019). A randomized education trial of spaced versus massed instruction to improve acquisition and retention of paediatric resuscitation skills in emergency medical service (EMS) providers. *Resuscitation*, 141, 73-80. <https://doi.org/10.1016/j.resuscitation.2019.06.010>
- *Pauwels, L., Swinnen, S. P., & Beets, I. A. (2014). Contextual interference in complex bimanual skill learning leads to better skill persistence. *PLoS ONE*, 9(6), e100906. <https://doi.org/10.1371/journal.pone.0100906>
- Peisah, C., Wijeratne, C., Waxman, B., & Vonau, M. (2014). Adaptive ageing surgeons. *ANZ Journal of Surgery*, 84(5), 311-315. <https://doi.org/10.1111/ans.12506>
- *Peltonen, V., Peltonen, L. M., Rantanen, M., Säämänen, J., Vääntinen, O., Koskela, J., Perkonoja, K., Salanterä, S., & Tommila, M. (2022). Randomized controlled trial comparing pit crew resuscitation model against standard advanced life support training. *Journal of the American College of Emergency Physicians Open*, 3(3), e12721. <https://doi.org/10.1002/emp2.12721>
- *Pemberton, J., Rambaran, M., & Cameron, B. H. (2013). Evaluating the long-term impact of the Trauma Team Training course in Guyana: an explanatory mixed-methods approach. *American Journal of Surgery*, 205(2), 119-124. <https://doi.org/10.1016/j.amjsurg.2012.08.004>
- Pendleton, I. P. (1986). *Factor contributing to loss of skills or knowledge gained in navy training courses* (Publication No. DP27138). [Doctoral dissertation, University of Southern California]. ProQuest Dissertations & Theses Global.
- *Peters, K. R., Smith, V., & Smith, C. T. (2007). Changes in sleep architecture following motor learning depend on initial skill level. *Journal of Cognitive Neuroscience*, 19(5), 817-829. <https://doi.org/10.1162/jocn.2007.19.5.817>
- Pieters, M. A., & Zaal, P. M. (2019). Training for Long-duration space missions: a literature review into skill retention and generalizability. *IFAC-PapersOnLine*, 52(19), 247-252. <https://doi.org/10.1016/j.ifacol.2019.12.099>

- *Pinto-Zipp, G. (1996). *Practice schedule and motor learning: Children versus adults* (Publication No. 9636015). [Doctoral dissertation, Teachers College, Columbia University]. ProQuest Dissertations & Theses Global.
- *Pollatou, E., Kioumourtzoglou, E., Agelousis, N., & Mavromatis, G. (1997). Contextual interference effects in learning novel motor skills. *Perceptual and Motor Skills*, 84(2), 487-496. <https://doi.org/10.2466/pms.1997.84.2.487>
- Polyany, M. (1967). *The tacit dimension*. London: Routledge & Kegan Paul.
- *Porte, M. C., Xeroulis, G., Reznick, R. K., & Dubrowski, A. (2007). Verbal feedback from an expert is more effective than self-accessed feedback about motion efficiency in learning new surgical skills. *American Journal of Surgery*, 193(1), 105-110. <https://doi.org/10.1016/j.amjsurg.2006.03.016>
- *Porter, C., Greenwood, D., Panchuk, D., & Pepping, G. J. (2020). Learner-adapted practice promotes skill transfer in unskilled adults learning the basketball set shot. *European Journal of Sport Science*, 20(1), 61-71. <https://doi.org/10.1080/17461391.2019.1611931>
- Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology*, 83(2, Pt.1), 304–308. <https://doi.org/10.1037/h0028558>
- Poulton, E. C. (1957). On the stimulus and response in pursuit tracking. *Journal of Experimental Psychology*, 53(3), 189–194. <https://doi.org/10.1037/h0043798>
- *Rafiee, S., & Dana, A. (2019). The effect of observing different information on learning the basketball jump shot. *Acta Gymnica*, 49(4), 164-173. <https://doi.org/10.5507/ag.2019.015>
- *Raisbeck, L. D., Regal, A., Diekfuss, J. A., Rhea, C. K., & Ward, P. (2015). Influence of practice schedules and attention on skill development and retention. *Human Movement Science*, 43, 100-106. <https://doi.org/10.1016/j.humov.2015.07.004>
- *Ramirez, A. G., Hu, Y. N., Kim, H., & Rasmussen, S. K. (2018). Long-Term Skills Retention Following a Randomized Prospective Trial on Adaptive Procedural Training. *Journal of Surgical Education*, 75(6), 1589-1597. <https://doi.org/10.1016/j.jsurg.2018.03.007>

- *Regal, A. B. (2013). *Skill acquisition and the influence of attentional focus and practice*. [Master's thesis, Michigan Technical University]. MTU Digital Commons. <https://doi.org/10.37099/mtu.dc.etsd/487>
- *Reis, J., Schambra, H. M., Cohen, L. G., Buch, E. R., Fritsch, B., Zarahn, E., Celnik, P. A., & Krakauer, J. W. (2009). Noninvasive cortical stimulation enhances motor skill acquisition over multiple days through an effect on consolidation. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(5), 1590-1595. <https://doi.org/10.1073/pnas.0805413106>
- *Rendell, M. A., Masters, R. S. W., Farrow, D., & Morris, T. (2011). An implicit basis for the retention benefits of random practice. *Journal of Motor Behavior*, *43*(1), 1-13. <https://doi.org/10.1080/00222895.2010.530304>
- *Riolo, L. (1997). Effects of modeling errors on the acquisition and retention of sterile hand washing task. *Perceptual and Motor Skills*, *84*(1), 19-26. <https://doi.org/10.2466/pms.1997.84.1.19>
- *Rivard, J. D., Vergis, A. S., Unger, B. J., Gillman, L. M., Hardy, K. M., & Park, J. (2015). The effect of blocked versus random task practice schedules on the acquisition and retention of surgical skills. *American Journal of Surgery*, *209*(1), 93-100. <https://doi.org/10.1016/j.amjsurg.2014.08.038>
- *Rodrigue, K. M., Kennedy, K. M., & Raz, N. (2005). Aging and longitudinal change in perceptual-motor skill acquisition in healthy adults. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *60*(4), P174-181. <https://doi.org/10.1093/geronb/60.4.p174>
- *Rodrigues, S. T., Ferracioli Mde, C., & Denardi, R. A. (2010). Learning a complex motor skill from video and point-light demonstrations. *Perceptual and Motor Skills*, *111*(2), 307-323. <https://doi.org/10.2466/05.11.23.24.25.Pms.111.5.307-323>
- *Rose, A. M., Czarnolewski, M. Y., Gragg, F. E., Austin, S. H., Ford, P., Doyle, J., & Hagman, J. D. (1985). *Acquisition and retention of soldiering skills* (Technical Report 671). Alexandria, VA: US Army Research Institute for Behavioral and Social Science. <https://apps.dtic.mil/sti/citations/tr/ADA160336>
- *Rosen, B. T., Uddin, P. Q., Harrington, A. R., Ault, B. W., & Ault, M. J. (2009). Does personalized vascular access training on a nonhuman tissue model allow for learning and retention of central line placement skills? Phase II of the Procedural

Patient Safety Initiative (PPSI-II). *Journal of Hospital Medicine*, 4(7), 423-429.
<https://doi.org/10.1002/jhm.571>

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Newbury Park, CA: SAGE Publications, Incorporated.

*Rosenthal, M. E., Ritter, E. M., Goova, M. T., Castellvi, A. O., Tesfay, S. T., Pimentel, E. A., Hartzler, R., & Scott, D. J. (2010). Proficiency-based fundamentals of laparoscopic surgery skills training results in durable performance improvement and a uniform certification pass rate. *Surgical Endoscopy*, 24(10), 2453-2457.
<https://doi.org/10.1007/s00464-010-0985-2>

Rowatt, W. C. & Shlechter, T. M. (1993). *A review and annotated bibliography of literature relevant to armor skill retention* (ARI Research Product 93-03). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. <https://apps.dtic.mil/sti/citations/ADA263407>

*Ruetzler, K., Roessler, B., Potura, L., Priemayr, A., Robak, O., Schuster, E., & Frass, M. (2011). Performance and skill retention of intubation by paramedics using seven different airway devices--a manikin study. *Resuscitation*, 82(5), 593-597.
<https://doi.org/10.1016/j.resuscitation.2011.01.008>

*de Ruijter, P. A., Biersteker, H. A., Biert, J., Van Goor, H., & Tan, E. C. (2014). Retention of first aid and basic life support skills in undergraduate medical students. *Medical Education Online*, 19(1), 24841.
<https://doi.org/10.3402/meo.v19.24841>

*Russell, D. M., & Newell, K. M. (2007). How persistent and general is the contextual interference effect? *Research Quarterly for Exercise and Sport*, 78(4), 318-327.
<https://doi.org/10.1080/02701367.2007.10599429>

*Ryan, E. D. (1962). Retention of stabilometer and pursuit rotor skills. *Research Quarterly of the American Association for Health, Physical Education, & Recreation*, 33(4), 593-598. <https://doi.org/10.1080/10671188.1962.10762112>

*Ryhanen, N. M. E. (2002). *Effects of serial and blocked practice on learning racquetball skills* (Publication No. 1410868). [Master's thesis, University of South Alabama]. ProQuest Dissertations & Theses Global.

Ryle, G. (1949/2000). *The concept of mind*. Chicago: University of Chicago Press.

- Sabol, M. A., & Wisher, R. A. (2001). Retention and reacquisition of military skills. *Military Operations Research*, 59-80.
- *Saiboon, I. M., Qamruddin, R. M., Jaafar, J. M., Bakar, A. A., Hamzah, F. A., Eng, H. S., & Robertson, C. E. (2016). Effectiveness of teaching automated external defibrillators use using a traditional classroom instruction versus self-instruction video in non-critical care nurses. *Saudi Medical Journal*, 37(4), 429-435. <https://doi.org/10.15537/smj.2016.4.14833>
- *Sale, M. V., & Kuzovina, A. (2022). Motor training is improved by concurrent application of slow oscillating transcranial alternating current stimulation to motor cortex. *BMC Neuroscience*, 23(1), 45. <https://doi.org/10.1186/s12868-022-00731-x>
- *Sall, D., Warm, E. J., Kinnear, B., Kelleher, M., Jandarov, R., & O'Toole, J. (2021). See one, do one, forget one: Early skill decay after paracentesis training. *Journal of General Internal Medicine*, 36(5), 1346-1351. <https://doi.org/10.1007/s11606-020-06242-x>
- *Sanders, W. R. (1999). *Digital procedural skill retention for selected M1A2 tank inter-vehicular information system (IVIS) tasks* (ARI Technical Report 1096). Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences. <https://apps.dtic.mil/sti/citations/ADA368212>
- *Sankar, J., Vijayakanthi, N., Sankar, M. J., & Dubey, N. (2013). Knowledge and skill retention of in-service versus preservice nursing professionals following an informal training program in pediatric cardiopulmonary resuscitation: a repeated-measures quasiexperimental study. *BioMed Research International*, 2013, 403415. <https://doi.org/10.1155/2013/403415>
- *Sankaranarayanan, G., Odlozil, C. A., Hasan, S. S., Shabbir, R., Qi, D., Turkseven, M., De, S., Funk, G., & Weddle, R. J. (2022). Training on a virtual reality cricothyroidotomy simulator improves skills and transfers to a simulated procedure. *Trauma Surgery & Acute Care Open*, 7(1), e000826. <https://doi.org/10.1136/tsaco-2021-000826>
- *Sant'Ana, G. M., Cavalini, W., Negrello, B., Bonin, E. A., Dimbarre, D., Claus, C., Loureiro, M. P., & Salvalaggio, P. R. (2017). Retention of laparoscopic skills in naive medical students who underwent short training. *Surgical Endoscopy*, 31(2), 937-944. <https://doi.org/10.1007/s00464-016-5063-y>

- *Saraç, L., & Ok, A. (2010). The effects of different instructional methods on students' acquisition and retention of cardiopulmonary resuscitation skills. *Resuscitation*, 81(5), 555-561. <https://doi.org/10.1016/j.resuscitation.2009.08.030>
- *Sarantinos, G. D. (1999). *Learning and retention adaptations of myoelectric activity during a novel multi-joint task*. [Master's Thesis, McGill University]. McGill University Repository. <https://escholarship.mcgill.ca/concern/theses/8s45qb94x>
- *Saucedo Marquez, C. M., Zhang, X., Swinnen, S. P., Meesen, R., & Wenderoth, N. (2013). Task-specific effect of transcranial direct current stimulation on motor learning. *Frontiers in Human Neuroscience*, 7, 333. <https://doi.org/10.3389/fnhum.2013.00333>
- *Scerbo, M. W., Britt, R. C., Montano, M., Kennedy, R. A., Prytz, E., & Stefanidis, D. (2017). Effects of a retention interval and refresher session on intracorporeal suturing and knot tying skill and mental workload. *Surgery*, 161(5), 1209-1214. <https://doi.org/10.1016/j.surg.2016.11.011>
- *Schendel, J., & Hagman, J. (1982). On sustaining procedural skills over a prolonged retention interval. *Journal of Applied Psychology*. <https://doi.org/10.1037/0021-9010.67.5.605>
- *Schmidt, M., Kemena, M., & Jaitner, T. (2021). Null Effects of Different Amounts of Task Variation in Both Contextual Interference and Differential Learning Paradigms. *Perceptual and Motor Skills*, 128(4), 1836-1850. <https://doi.org/10.1177/00315125211022302>
- *Schott, C. K., LoPresti, C. M., Boyd, J. S., Core, M., Haro, E. K., Mader, M. J., Pascual, S., Finley, E. P., Lucas, B. P., Colon-Molero, A., Restrepo, M. I., Pugh, J., & Soni, N. J. (2021). Retention of point-of-care ultrasound skills among practicing physicians: Findings of the VA national POCUS training program. *American Journal of Medicine*, 134(3), 391-399.e398. <https://doi.org/10.1016/j.amjmed.2020.08.008>
- *Seifert, L. B., Herrera-Vizcaino, C., Herguth, P., Sterz, J., & Sader, R. (2020). Comparison of different feedback modalities for the training of procedural skills in Oral and maxillofacial surgery: a blinded, randomized and controlled study. *BMC Medical Education*, 20(1), 330. <https://doi.org/10.1186/s12909-020-02222-1>
- *Seifert, L. B., Schnurr, B., Stefanescu, M. C., Sader, R., Ruessler, M., & Sterz, J. (2020). Comparing video-based versions of Halsted's 'see one, do one' and

Peyton's '4-step approach' for teaching surgical skills: A randomized controlled trial. *BMC Medical Education*, 20(1), 194. <https://doi.org/10.1186/s12909-020-02105-5>

- *Sekiya, H., Magill, R. A., Sidaway, B., & Anderson, D. I. (1994). The contextual interference effect for skill variations from the same and different generalized motor programs. *Research Quarterly for Exercise and Sport*, 65(4), 330-338. <https://doi.org/10.1080/02701367.1994.10607637>

- *Shafizadeh, M., Platt, G. K., & Bahram, A. (2013). Effects of focus of attention and type of practice on learning and self-efficacy in dart throwing. *Perceptual and Motor Skills*, 117(1), 182-192. <https://doi.org/10.2466/25.30.PMS.117x12z5>

- *Shanks, D., Brydges, R., den Brok, W., Nair, P., & Hatala, R. (2013). Are two heads better than one? Comparing dyad and self-regulated learning in simulation training. *Medical Education*, 7(12), 1215-1222. <https://doi.org/10.1111/medu.12284>

- *Shappell, R. T. (1977). *The effect of three types of mental practice on performance and retention scores of a maze tracking task* (Publication No. 7724805). [Doctoral dissertation, Florida State University]. ProQuest Dissertations & Theses Global.

- *Shaver, S. L., Yamada, N., & Hofmeister, E. H. (2020). Retention of basic suturing skills with brief or extended practice in veterinary students. *Veterinary Surgery*, 49(6), 1239-1245. <https://doi.org/10.1111/vsu.13439>

- *Shea, C. H., Lai, Q., Black, C., & Park, J. H. (2000). Spacing practice sessions across days benefits the learning of motor skills. *Human Movement Science*, 19(5), 737-760. [https://doi.org/10.1016/S0167-9457\(00\)00021-X](https://doi.org/10.1016/S0167-9457(00)00021-X)

- *Sheedy, J. (1986). *Mental rehearsal and skill acquisition in sports*. [Master's thesis, University of Wollongong]. University of Wollongong Repository. <https://ro.uow.edu.au/theses/2251/>

- *Sheth, B. R., Janvelyan, D., & Khan, M. (2008). Practice makes imperfect: Restorative effects of sleep on motor learning. *PLoS ONE*, 3(9), e3190. <https://doi.org/10.1371/journal.pone.0003190>

- *Shewokis, P. A. (1997). Is the contextual interference effect generalizable to computer games? *Perceptual and Motor Skills*, 84(1), 3-15.
<https://doi.org/10.2466/pms.1997.84.1.3>
- *Shewokis, P. A., Krane, V., Snow, J., & Greenleaf, C. (2001). Does trait cognitive anxiety influence the learning of perceptual-motor skills in a contextual interference paradigm? *Journal of Human Movement Studies*, 41(3), 225-245.
- *Shippey, S. H., Chen, T. L., Chou, B., Knoepp, L. R., Bowen, C. W., & Handa, V. L. (2011). Teaching subcuticular suturing to medical students: video versus expert instructor feedback. *Journal of Surgical Education*, 68(5), 397-402.
<https://doi.org/10.1016/j.jsurg.2011.04.006>
- *Shoenfelt, E. L., Snyder, L. A., Maue, A. E., McDowell, C. P., & Woolard, C. D. (2002). Comparison of constant and variable practice conditions on free-throw shooting. *Perceptual and Motor Skills*, 94(3 Pt 2), 1113-1123.
<https://doi.org/10.2466/pms.2002.94.3c.1113>
- *Sidaway, B., & Hand, M. J. (1993). Frequency of modeling effects on the acquisition and retention of a motor skill. *Research Quarterly for Exercise and Sport*, 64(1), 122-126. <https://doi.org/10.1080/02701367.1993.10608786>
- *Siegel, J. (1963). *Post-training performance as a function of schedule and specificity of knowledge of results during training* (Publication No. 6400270). [Doctoral dissertation, New York University]. ProQuest Dissertations & Theses Global.
- *Sigrist, R., Rauter, G., Marchal-Crespo, L., Riener, R., & Wolf, P. (2015). Sonification and haptic feedback in addition to visual feedback enhances complex motor task learning. *Experimental Brain Research*, 233(3), 909-925.
<https://doi.org/10.1007/s00221-014-4167-7>
- *Simon, D. A. (2007). Contextual interference effects with two tasks. *Perceptual and Motor Skills*, 105(1), 177-183. <https://doi.org/10.2466/PMS.105.1.177-183>
- *Simon, D. A., & Bjork, R. A. (2002). Models of performance in learning multisegment movement tasks: Consequences for acquisition, retention, and judgments of learning. *Journal of Experimental Psychology: Applied*, 8(4), 222-232.
<https://doi.org/10.1037/1076-898X.8.4.222>

- *Singer, R. N. (1965). Massed and distributed practice effects on the acquisition and retention of a novel basketball skill. *Research Quarterly, American Association for Health, Physical Education and Recreation*, 36(1), 68-77.
<https://doi.org/10.1080/10671188.1965.10614658>
- *Sinha, P., Hogle, N. J., & Fowler, D. L. (2008). Do the laparoscopic skills of trainees deteriorate over time? *Surgical Endoscopy*, 22, 2018-2025.
<https://doi.org/10.1007/s00464-008-9929-5>
- Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. *Psychological Review*, 82(4), 225–260. <https://doi.org/10.1037/h0076770>
- Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis*. London: SAGE Publications, Ltd.
- *Skinner, A. D. (2013). *Retention and retraining of independent and integrated cognitive and psychomotor skills related to laparoscopic surgery*. [Doctoral dissertation, The Catholic University of America].
<https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=3a6c0d8494ec12c578ef7f8f447ac02e26c32dbe>
- *Smit, D., Spruit, E., Dankelman, J., Tuijthof, G., Hamming, J., & Horeman, T. (2017). Improving training of laparoscopic tissue manipulation skills using various visual force feedback types. *Surgical Endoscopy*, 31, 299-308.
<https://doi.org/10.1007/s00464-016-4972-0>
- *Smith, C. C., Huang, G. C., Newman, L. R., Clardy, P. F., Feller-Kopman, D., Cho, M., Ennacheril, T., & Schwartzstein, R. M. (2010). Simulation training and its effect on long-term resident performance in central venous catheterization. *Simulation In Healthcare*, 5(3), 146-151. <https://doi.org/10.1097/SIH.0b013e3181dd9672>
- *Smith, J., Doody, K., & Veitch, B. (2019). Being prepared for emergencies: a virtual environment experiment on the retention and maintenance of egress skills. *WMU Journal of Maritime Affairs*, 18(3), 425-449. <https://doi.org/10.1007/s13437-019-00174-y>
- *Smith, P. J. K. (2002). Applying contextual interference to snowboarding skills. *Perceptual and Motor Skills*, 95(3 PART 1), 999-1005.
<https://doi.org/10.2466/pms.2002.95.3.999>

- *Smith, S. J., Farra, S., Ulrich, D. L., Hodgson, E., Nicely, S., & Matcham, W. (2016). Learning and retention using virtual reality in a decontamination simulation. *Nursing Education Perspectives*, 37(4), 210-214. <https://doi.org/10.1097/01.Nep.0000000000000035>
- *So, J. C., Proctor, R. W., Dunston, P. S., & Wang, X. (2013). Better retention of skill operating a simulated hydraulic excavator after part-task than after whole-task training. *Human Factors*, 55(2), 449-460. <https://doi.org/10.1177/0018720812454292>
- *de Souza Nunes, M. E., Correa, U. C., de Souza, M. G. T. X., & Santos, S. (2021). Descriptive versus prescriptive feedback in the learning of golf putting by older persons. *International Journal of Sport and Exercise Psychology*, 19(4), 709-721. <https://doi.org/10.1080/1612197X.2020.1717579>
- *de Souza Nunes, M. E., Correa, U. C., de Souza, M. G. T. X., Basso, L., Coelho, D. B., & Santos, S. (2019). No improvement on the learning of golf putting by older persons with self-controlled knowledge of performance. *Journal of Aging and Physical Activity*, 27(3), 300-308. <https://doi.org/10.1123/japa.2018-0053>
- *de Souza Nunes, M. G. T. X., Nunes, M. E. S., Corrêa, U. C., & Dos Santos, S. (2015). The contextual interference effect on sport-specific motor learning in older adults. *Human Movement*, 16(3), 112-118. <https://doi.org/10.1515/humo-2015-0036>
- *Spittle, M., & Kremer, P. (2010). Mental practice and the retention of motor learning: a pilot study. *Perceptual and Motor Skills*, 110(3 Pt 1), 888-896. <https://doi.org/10.2466/pms.110.3.888-896>
- *Spooner, B. B., Fallaha, J. F., Kocierz, L., Smith, C. M., Smith, S. C., & Perkins, G. D. (2007). An evaluation of objective feedback in basic life support (BLS) training. *Resuscitation*, 73(3), 417-424. <https://doi.org/10.1016/j.resuscitation.2006.10.017>
- *St. Germain, L., McKay, B., Poskus, A., Williams, A., Leshchyshe, O., Feldman, S., Cashaback, J. G. A., & Carter, M. J. (2022). Exercising choice over feedback schedules during practice is not advantageous for motor learning. *Psychonomic Bulletin and Review*. <https://doi.org/10.3758/s13423-022-02170-5>
- Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, 5(1), 60-78. <https://doi.org/10.1002/jrsm.1095>

- Stansfield, T., & Tai, N. (2021). Surgical skill decay in the contingency era. *BMJ Military Health*, *167*(5), 300-301. <https://doi.org/10.1136/bmjmilitary-2021-001921>
- *Starodub, R., Abella, B. S., Hoyt-Brennan, A. M., Leary, M., Mancini, M. E., Chittams, J., & Riegel, B. (2020). A comparative study of video lecture versus video lecture and high fidelity simulation for training nurses on the delivery of targeted temperature management after cardiac arrest. *International Emergency Nursing*, *49*, 100829. <https://doi.org/10.1016/j.ienj.2019.100829>
- *Statton, M. A., Encarnacion, M., Celnik, P., & Bastian, A. J. (2015). A single bout of moderate aerobic exercise improves motor skill acquisition. *PLoS ONE*, *10*(10). <https://doi.org/10.1371/journal.pone.0141393>
- *Stefanidis, D., Korndorffer, J. R., Jr., Markley, S., Sierra, R., & Scott, D. J. (2006). Proficiency maintenance: impact of ongoing simulator training on laparoscopic skill retention. *Journal of the American College of Surgeons*, *202*(4), 599-603. <https://doi.org/10.1016/j.jamcollsurg.2005.12.018>
- Sterne, J. A. C, Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *Journal of Clinical Epidemiology*, *54*(10), 1046–1055. [https://doi.org/10.1016/S0895-4356\(01\)00377-8](https://doi.org/10.1016/S0895-4356(01)00377-8)
- Stross, J. K. (1983). Maintaining competency in advanced cardiac life support skills. *JAMA*, *249*(24), 3339-3341. <https://doi.org/10.1001/jama.1983.03330480045028>
- *Suddon, F. H., & Lavery, J. J. (1962). The effect of amount of training on retention of a simple motor skill with 0-and 5-trial delays of knowledge of results. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *16*(4), 312. <https://doi.org/10.1037/h0083258>
- *Suebnuarn, S., Haddawy, P., Rhienmora, P., Jittimane, P., & Viratket, P. (2010). Augmented kinematic feedback from haptic virtual reality for dental skill acquisition. *Journal of Dental Education*, *74*(12), 1357-1366. <https://doi.org/10.1002/j.0022-0337.2010.74.12.tb05011.x>
- *Suet, G., Blanie, A., de Montblanc, J., Roulleau, P., & Benhamou, D. (2020). External cardiac massage training of medical students: A randomized comparison of two feedback Methods to Standard Training. *Journal of Emergency Medicine*, *59*(2), 270-277. <https://doi.org/10.1016/j.jemermed.2020.04.058>

- Sullivan, A., Elshenawy, S., Ades, A., & Sawyer, T. (2019). Acquiring and maintaining technical skills using simulation: initial, maintenance, booster, and refresher training. *Cureus, 11*(9), e5729. <https://doi.org/10.7759/cureus.5729>
- *Surburg, P. R. (1976). Aging and effect of physical-mental practice upon acquisition and retention of a motor skill. *Journal of Gerontology, 31*(1), 64-67. <https://doi.org/10.1093/geronj/31.1.64>
- *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*(6), 751-762. <https://doi.org/10.1177/001872088803000611>
- Swift, E. J. (1905). Memory of a complex skillful act. *The American Journal of Psychology, 16*(1), 131-133. <https://doi.org/10.2307/1412230>
- Swift, E. J. (1906). Memory of skillful movements. *Psychological Bulletin, 3*(6), 185–187. <https://doi.org/10.1037/h0074656>
- Swift, E. J. (1910). Relearning a skillful act. An experimental study in neuro-muscular memory. *Psychological Bulletin, 7*(1), 17–19. <https://doi.org/10.1037/h0071771>
- Symmons, M., & Mulvihill, C. (2011, June). A simulator comparison of riding performance between new, returned and continuing motorcycle riders. In *Driving Assesment Conference* (Vol. 6, No. 2011). University of Iowa. <https://doi.org/10.17077/drivingassessment.1443>
- *Teishima, J., Hattori, M., Inoue, S., Ikeda, K., Hieda, K., Ohara, S., Egi, H., Ohdan, H., & Matsubara, A. (2014). Retention of robot-assisted surgical skills in urological surgeons acquired using Mimic dV-Trainer. *Canadian Urological Association Journal, 8*(7-8), e493-e497. <https://doi.org/10.5489/cuaj.1993>
- *Thomas, R., Flindtgaard, M., Skriver, K., Geertsen, S. S., Christiansen, L., Johnsen, L. K., Busk, D. V. P., Bojsen-Moller, E., Madsen, M. J., Ritz, C., Roig, M., & Lundbye-Jensen, J. (2017). Acute exercise and motor memory consolidation: Does exercise type play a role? *Scandanavian Journal of Medicine & Science in Sports, 27*(11), 1523-1532. <https://doi.org/10.1111/sms.12791>
- *Thomas, R., Johnsen, L. K., Geertsen, S. S., Christiansen, L., Ritz, C., Roig, M., & Lundbye-Jensen, J. (2016). Acute exercise and motor memory consolidation: The

role of exercise intensity. *PLoS ONE*, *11*(7), e0159589.
<https://doi.org/10.1371/journal.pone.0159589>

- *Thomas, S. M., Burch, W., Kuehnle, S. E., Flood, R. G., Scalzo, A. J., & Gerard, J. M. (2013). Simulation training for pediatric residents on central venous catheter placement: a pilot study. *Pediatric Critical Care Medicine*, *14*(9), e416-423.
<https://doi.org/10.1097/PCC.0b013e31829f5eda>
- *Thomaschewski, M., Esnaashari, H., Höfer, A., Renner, L., Benecke, C., Zimmermann, M., Keck, T., & Laubert, T. (2019). Video tutorials increase precision in minimally invasive surgery training - A prospective randomised trial and follow-up study. *Zentralblatt für Chirurgie - Zeitschrift für Allgemeine, Viszeral- und Gefasschirurgie*, *144*(2), 153-161. <https://doi.org/10.1055/a-0638-8295>
- *Thompson, T. J., Moray, J. C., & Smith, S. (1981). *Basic rifle marksmanship skill retention: Implications for retention research* (Research Report 1326). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. <https://apps.dtic.mil/sti/citations/tr/ADA134017>
- *Tomporowski, P. D., & Pendleton, D. M. (2018). Effects of the timing of acute exercise and movement complexity on young adults' psychomotor learning. *Journal of Sport and Exercise Psychology*, *40*(5), 240-248.
<https://doi.org/10.1123/jsep.2017-0289>
- *Tong, C., & Flanagan, J. R. (2003). Task-specific internal models for kinematic transformations. *Journal of Neurophysiology*, *90*(2), 578-585.
<https://doi.org/10.1152/jn.01087.2002>
- *Torkington, J., Smith, S. G., Rees, B., & Darzi, A. (2001). The role of the basic surgical skills course in the acquisition and retention of laparoscopic skill. *Surgical Endoscopy*, *15*(10), 1071-1075. <https://doi.org/10.1007/s004640000183>
- *Toy, S., McKay, R., Eilert, R., & Sandall, J. (2019). Effective and feasible simulation-based procedural training for medical students: Instructional video-guided deliberate practice versus training with expert feedback. *Medical Science Educator*, *29*(1), 35-39. <https://doi.org/10.1007/s40670-018-00643-8>
- *Truong, C., Hilt, P. M., Bouguila, F., Bove, M., Lebon, F., Papaxanthis, C., & Ruffino, C. (2022). Time-of-day effects on skill acquisition and consolidation after physical and mental practices. *Scientific Reports*, *12*(1), 5933.
<https://doi.org/10.1038/s41598-022-09749-x>

- *Tse, A. C. Y., Wong, T. W. L., & Masters, R. S. W. (2017). Examining motor learning in older adults using analogy instruction. *Psychology of Sport and Exercise*, 28, 78-84. <https://doi.org/10.1016/j.psychsport.2016.10.005>
- *Udani, A. D., Harrison, T. K., Mariano, E. R., Derby, R., Kan, J., Ganaway, T., Shum, C., Gaba, D. M., Tanaka, P., Kou, A., Howard, S. K., & Group, A. R. (2016). Comparative-effectiveness of simulation-based deliberate practice versus self-guided practice on resident anesthesiologists' acquisition of ultrasound-guided regional anesthesia skills. *Regional Anesthesia and Pain Medicine*, 41(2), 151-157. <https://doi.org/10.1097/AAP.0000000000000361>
- Ulleberg, P., Bjørnskau, T., & Fostervold, K. I. (2022). Does age matter? Examining age-dependent differences in at-fault collisions after attending a refresher course for older drivers. *Transportation Research Part F: Traffic Psychology and Behaviour*, 87, 379-390. <https://doi.org/10.1016/j.trf.2022.04.016>
- *Vallabhajosula, S., Judkins, T. N., Mukherjee, M., Suh, I. H., Oleynikov, D., & Siu, K. C. (2013). Skills learning in robot-assisted surgery is benefited by task-specific augmented feedback. *Surgical Innovation*, 20(6), 639-647. <https://doi.org/10.1177/1553350613484590>
- *Van Bruwaene, S., De Win, G., & Miserez, M. (2009). How much do we need experts during laparoscopic suturing training? *Surgical Endoscopy*, 23(12), 2755-2761. <https://doi.org/10.1007/s00464-009-0498-z>
- *Van Hove, C., Perry, K. A., Spight, D. H., Wheeler-McInville, K., Diggs, B. S., Sheppard, B. C., Jobe, B. A., & O'Rourke, R. W. (2008). Predictors of technical skill acquisition among resident trainees in a laparoscopic skills education program. *World Journal of Surgery*, 32(9), 1917-1921. <https://doi.org/10.1007/s00268-008-9643-4>
- *Veldman, M. P., Maurits, N. M., Mantini, D., & Hortobágyi, T. (2021). Age-dependent modulation of motor network connectivity for skill acquisition, consolidation and interlimb transfer after motor practice. *Clinical Neurophysiology*, 132(8), 1790-1801. <https://doi.org/10.1016/j.clinph.2021.03.051>
- *Veldman, M. P., Maurits, N. M., Zijdwind, I., Maffiuletti, N. A., van Middelkoop, S., Mizelle, J. C., & Hortobágyi, T. (2018). Somatosensory electrical stimulation improves skill acquisition, consolidation, and transfer by increasing sensorimotor activity and connectivity. *Journal of Neurophysiology*, 120(1), 281-290. <https://doi.org/10.1152/jn.00860.2017>

- *Veldman, M. P., Zijdwind, I., Maffiuletti, N. A., & Hortobágyi, T. (2016). Motor skill acquisition and retention after somatosensory electrical stimulation in healthy humans. *Frontiers in Human Neuroscience*, *10*.
<https://doi.org/10.3389/fnhum.2016.00115>
- *Verdolinimarston, K., & Balota, D. A. (1994). Role of elaborative and perceptual integrative processes in perceptual-motor performance. *Journal of Experimental Psychology-Learning Memory and Cognition*, *20*(3), 739-749.
<https://doi.org/10.1037/0278-7393.20.3.739>
- Villado, A. J., Day, E. A., Arthur Jr, W., Boatman, P. R., Kowollik, V., Bhupatkar, A., & Bennett Jr, W. (2013). Complex command-and-control simulation task performance following periods of nonuse. In *Individual and team skill decay* (pp. 77-91). Routledge.
- *Vine, S. J., Chaytor, R. J., McGrath, J. S., Masters, R. S. W., & Wilson, M. R. (2013). Gaze training improves the retention and transfer of laparoscopic technical skills in novices. *Surgical Endoscopy*, *27*(9), 3205-3213.
<https://doi.org/10.1007/s00464-013-2893-8>
- *Voscopoulos, C., Barker, T., Listwa, T., Nelson, S., Pozner, C., Liu, X., Zane, R., & Antoine, J. A. (2013). A comparison of the speed, success rate, and retention of rescue airway devices placed by first-responder emergency medical technicians: a high-fidelity human patient simulation study. *Journal of Emergency Medicine*, *44*(4), 784-789. <https://doi.org/10.1016/j.jemermed.2012.07.045>
- *Wang, D., Li, T., Yang, G., & Zhang, Y. (2017). Effects of concurrent and delayed visual feedback on motor memory consolidation. *IEEE Trans Haptics*, *10*(3), 350-357. <https://doi.org/10.1109/toh.2017.2672549>
- *Wang, P., Infurna, F. J., & Schaefer, S. Y. (2020). Predicting motor skill learning in older adults using visuospatial performance. *Journal of Motor Learning and Development*, *8*(1), 38-51. <https://doi.org/10.1123/jmld.2018-0017>
- Wang, X. (2010). *Factors influencing knowledge and skill decay in organizational training: A meta-analysis* (Publication No. 3404994). [Doctoral Dissertation, University of Oklahoma]. ProQuest Dissertations & Theses Global.
- Wang, X, Day, E. A., Kowollik, V., Schuelke M. J., Hughes M. G. (2013). Factors influencing knowledge and skill decay after training. In *Individual and Team Skill Decay: The Science and Implications for Practice*, ed. W Arthur Jr., EA Day, W

Bennett Jr., AM Portrey, pp. 68–116. New York: New York: Routledge/Taylor & Francis/Psychology Press.

- *Warner-Codish, K. A. (2018). *Arm ergometry exercise intensity interaction with motor memory* [Doctoral dissertation, Texas Women's University]. TWU Repository. <https://hdl.handle.net/11274/10735>

- *Wayne, D. B., Siddall, V. J., Butter, J., Fudala, M. J., Wade, L. D., Feinglass, J., & McGaghie, W. C. (2006). A longitudinal study of internal medicine residents' retention of advanced cardiac life support skills. *Academic Medicine*, *81*(10), S9-S12.

- *Wayne, D. B., Siddall, V. J., Butter, J., Fudala, M. J., Wade, L. D., Feinglass, J., & McGaghie, W. C. (2006). A longitudinal study of internal medicine residents' retention of advanced cardiac life support skills. *Academic Medicine*, *81*(10 Suppl), S9-s12. <https://doi.org/10.1097/00001888-200610001-00004>

- *Weeks, D. L., Wallace, S. A., & Anderson, D. I. (2003). Training with an upper-limb prosthetic simulator to enhance transfer of skill across limbs. *Archives of Physical Medicine and Rehabilitation*, *84*(3), 437-443. <https://doi.org/10.1053/apmr.2003.50014>

- *Wehlmann, J. A., & Wulf, G. (2021). Bullseye: Effects of autonomy support and enhanced expectancies on dart throwing. *International Journal of Sports Science & Coaching*, *16*(2), 317-323. <https://doi.org/10.1177/1747954120967788>

- *Weis, J. J., Farr, D., Abdelfattah, K. R., Hogg, D., & Scott, D. J. (2019). A proficiency-based surgical boot camp May not provide trainees with a durable foundation in fundamental surgical skills. *American Journal of Surgery*, *217*(2), 244-249. <https://doi.org/10.1016/j.amjsurg.2018.07.040>

- *Welsher, A., & Grierson, L. E. M. (2017). Enhancing technical skill learning through interleaved mixed-model observational practice. *Advances in Health Sciences Education*, *22*(5), 1201-1211. <https://doi.org/10.1007/s10459-017-9759-0>

- *Werman, H. A., Brown, C. G., Keseg, D. R., Glimcher, M., Schumacher, C., & Shaner, S. (1990). Retention of basic trauma life support skills. *Prehospital and Disaster Medicine*, *5*(2), 137-144. <https://doi.org/10.1017/S1049023X00026716>

- Wetzel, S. K., Konoske, P. J., & Montague, W. E. (1983). *Estimating skill degradation for aviation antisubmarine warfare operators (AWs): Loss of skill and knowledge following training* (Tech. Rep. No. NPRDC TR-83-31). San Diego: Navy Personnel Research and Development Center.
- Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica, 41*, 67-85.
- *Wiedenbeck, S. (1999). The use of icons and labels in an end user application program: An empirical study of learning and retention. *Behaviour & Information Technology, 18*(2), 68-82. <https://doi.org/10.1080/014492999119129>
- *Wiegardt, P. A. (1997). *The effect of visual imagery perspective on the learning, retention, and transfer of a complex field hockey skill* (Publication No. 9820301). [Doctoral dissertation, University of Virginia]. ProQuest Dissertations & Theses Global.
- *Wierinck, E., Puttemans, V., & van Steenberghe, D. (2006). Effect of tutorial input in addition to augmented feedback on manual dexterity training and its retention. *European Journal of Dental Education, 10*(1), 24-31. <https://doi.org/10.1111/j.1600-0579.2006.00392.x>
- *Wierinck, E., Puttemans, V., Swinnen, S., & van Steenberghe, D. (2005). Effect of augmented visual feedback from a virtual reality simulation system on manual dexterity training. *European Journal of Dental Education, 9*(1), 10-16. <https://doi.org/10.1111/j.1600-0579.2004.00351.x>
- *Wierinck, E. R., Puttemans, V., Swinnen, S. P., & van Steenberghe, D. (2007). Expert performance on a virtual reality simulation system. *Journal of Dental Education, 71*(6), 759-766. <https://doi.org/10.1002/j.0022-0337.2007.71.6.tb04332.x>
- *Wik, L., Myklebust, H., Auestad, B. H., & Steen, P. A. (2005). Twelve-month retention of CPR skills with automatic correcting verbal feedback. *Resuscitation, 66*(1), 27-30. <https://doi.org/10.1016/j.resuscitation.2004.12.022>
- *Williams, C. K., Tremblay, L., & Carnahan, H. (2016). It pays to go off-track: Practicing with error-augmenting haptic feedback facilitates learning of a curve-tracing task. *Frontiers in Psychology, 7*. <https://doi.org/10.3389/fpsyg.2016.02010>

- Willingham, D. B., Nissen, M. J., & Bullemer, P. (1989). On the development of procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(6), 1047–1060. [https://doi.org/10.1016/0001-6918\(77\)90012-9](https://doi.org/10.1016/0001-6918(77)90012-9)
- *Wilson, C. A., Davidson, J., Chahine, S., Chan, E. P., Stringer, L., Quantz, M. A., Saklofske, D. H., & Wang, P. Z. T. (2022). What is transferred and how much is retained? A simulation study of complex surgical skills. *Journal of Surgical Research*, 280, 411-420. <https://doi.org/10.1016/j.jss.2022.07.040>
- Wisher, R. A., Sabol, M. A., Ellis, J., & Ellis, K. (1999). *Staying sharp: Retention of military knowledge and skills* (U.S. Army Research Institute Special Report 39). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. <https://apps.dtic.mil/sti/pdfs/ADA366825.pdf>
- Wisher, R. A., Sabol, M. A., Hillel, H., & Kern, R. P. (1991). *Individual ready reserve (IRR) call-up: Skill decay*. (ARI Research Rep. No. 1595). Alexandria, VA: U.S. Army Research Institute.
- *Wohldmann, E. L. (2006). *Pushing the limits of imagination: The effectiveness of motor imagery for acquiring and maintaining a sequential motor skill* (Publication No. 3207743). [Doctoral dissertation, University of Colorado at Boulder]. ProQuest Dissertations & Theses Global.
- Wood, R. E. (1986). Task complexity: Definition of the construct. *Organizational Behavior and Human Decision Processes*, 37(1), 60-82. [https://doi.org/10.1016/0749-5978\(86\)90044-0](https://doi.org/10.1016/0749-5978(86)90044-0)
- *Wu, V., & Beyea, J. A. (2017). Evaluation of a web-based module and an otoscopy simulator in teaching ear disease. *Otolaryngology-Head and Neck Surgery*, 156(2), 272-277. <https://doi.org/10.1177/0194599816677697>
- *Wu, Y. H., Truglio, T. S., Zatsiorsky, V. M., & Latash, M. L. (2015). Learning to combine high variability with high precision: Lack of transfer to a different task. *Journal of Motor Behavior*, 47(2), 153-165. <https://doi.org/10.1080/00222895.2014.961892>
- *Wulf, G., & Adams, N. (2014). Small choices can enhance balance learning. *Human Movement Science*, 38, 235-240. <https://doi.org/10.1016/j.humov.2014.10.007>

- *Wulf, G., Chiviawsky, S., & Drews, R. (2015). External focus and autonomy support: Two important factors in motor learning have additive benefits. *Human Movement Science, 40*, 176-184. <https://doi.org/10.1016/j.humov.2014.11.015>
- *Wulf, G., Iwatsuki, T., Machin, B., Kellogg, J., Copeland, C., & Lewthwaite, R. (2018). Lassoing skill through learner choice. *Journal of Motor Behavior, 50*(3), 285-292. <https://doi.org/10.1080/00222895.2017.1341378>
- *Wynne, G. A. (1993). *Evaluation of different approaches to teaching basic resuscitation* (Publication No. U066101). [Master's thesis, University of London, Royal Free Hospital School of Medicine]. ProQuest Dissertations & Theses Global.
- *Xeroulis, G. J., Park, J., Moulton, C. A., Reznick, R. K., LeBlanc, V., & Dubrowski, A. (2007). Teaching suturing and knot-tying skills to medical students: A randomized controlled study comparing computer-based video instruction and (concurrent and summary) expert feedback. *Surgery, 141*(4), 442-449. <https://doi.org/10.1016/j.surg.2006.09.012>
- Xu, F. F., Ji, L., Shi, B., Du, J., Neubig, G., Bisk, Y., & Duan, N. (2020). A benchmark for structured procedural knowledge extraction from cooking videos. *arXiv preprint arXiv:2005.00706*. <https://doi.org/10.48550/arXiv.2005.00706>
- *Yamamoto, R., Clanton, D., Willis, R. E., Jonas, R. B., & Cestero, R. F. (2018). Rapid decay of transthoracic echocardiography skills at 1 month: A prospective observational study. *Journal of Surgical Education, 75*(2), 503-509. <https://doi.org/10.1016/j.jsurg.2017.07.011>
- *Yao, W. X., Fischman, M. G., & Wang, Y. T. (1994). Motor skill acquisition and retention as a function of average feedback, summary feedback, and performance variability. *Journal of Motor Behavior, 26*(3), 273-282. <https://doi.org/10.1080/00222895.1994.9941683>
- *Yechiam, E., Erev, I., Yehene, V., & Gopher, D. (2003). Melioration and the Transition from Touch-Typing Training to Everyday Use. *Human Factors, 45*(4), 671-684. <https://doi.org/10.1518/hfes.45.4.671.27085>
- *Yeo, C. T., Davison, C., Ungi, T., Holden, M., Fichtinger, G., & McGraw, R. (2015). Examination of learning trajectories for simulated lumbar puncture training using hand motion analysis. *Academic Emergency Medicine, 22*(10), 1187-1195. <https://doi.org/10.1111/acem.12753>

- *Zhang, K. J., Zhou, H., Guo, H., Li, W., Yang, Z., Liu, R., Qin, S., Xie, X., & Tian, J. (2022). Learning and Short-Term Retention of Simulation-Based Arthroscopic Skills. *Journal of Surgical Education*. <https://doi.org/10.1016/j.jsurg.2022.08.020>
- *Zheng, B., Fu, B., Al-Tayeb, T. A., Hao, Y. F., & Qayumi, A. K. (2014). Mastering instruments before operating on a patient: the role of simulation training in tool use skills. *Surgical Innovation*, *21*(6), 637-642. <https://doi.org/10.1177/1553350614532533>
- *Zhu, F. F., Poolton, J. M., Maxwell, J. P., Fan, J. K. M., Leung, G. K. K., & Masters, R. S. W. (2014). Refining the continuous tracking paradigm to investigate implicit motor learning. *Experimental Psychology*, *61*(3), 196-204. <https://doi.org/10.1027/1618-3169/a000239>
- *Zhu, T. S., Godse, N., Clayburgh, D. R., & Duvvuri, U. (2022). Assessing the learning curve associated with a novel flexible robot in the pre-clinical and clinical setting. *Surg Endosc*, *36*(2), 1563-1572. <https://doi.org/10.1007/s00464-021-08445-7>
- *Ziv, G., Lidor, R., Elbaz, L., & Lavie, M. (2020). Preference-Performance Dissociation in Golf Putting. *Frontiers in Psychology*, *11*, 102. <https://doi.org/10.3389/fpsyg.2020.00102>
- *Ziv, G., Lidor, R., & Lavie, M. (2021). Enhanced expectancies in golf putting – A replication study with increased ecological validity. *International Journal of Sport and Exercise Psychology*, *19*(1), 1-12. <https://doi.org/10.1080/1612197X.2019.1637362>
- *Zondervan, D. K., Duarte, J. E., Rowe, J. B., & Reinkensmeyer, D. J. (2014). Time flies when you are in a groove: Using entrainment to mechanical resonance to teach a desired movement distorts the perception of the movement's timing. *Experimental Brain Research*, *232*(3), 1057-1070. <https://doi.org/10.1007/s00221-013-3819-3>
- Zwetsloot, P. P., Van Der Naald, M., Sena, E. S., Howells, D. W., IntHout, J., De Groot, J. A., ... & Wever, K. E. (2017). Standardized mean differences cause funnel plot distortion in publication bias assessments. *Elife*, *6*, e24260. <https://doi.org/10.7554/eLife.24260>