REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
xx-09-2012	Technical	September 2012

4. TITLE AND SUBTITLE
Examine the Impact of Training Duration on Retention

5a. CONTRACT NUMBER
H92222-10-D-0017/0007

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

SWA Consulting, Inc.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

SWA Consulting Inc.
311 S. Harrington St.
Suite 200
Raleigh, NC 27603

8. PERFORMING ORGANIZATION REPORT NUMBER

2012010634

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

HQ USSOCOM, Special Operations Forces Language Office
Attn: FMD-LDEMO-TL
7701 Tampa Point Blvd.
MacDill AFB, FL 33621-5323

10. SPONSOR/MONITOR'S ACRONYM(S)

SOFLO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

A. Approved for public release; distribution is unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The purpose of this report is to provide Special Operations Forces (SOF) language program administrators with evidence-based conclusions and recommendations pertaining to how initial acquisition training (IAT) duration impacts SOF operators' ability to retain their language capability, specifically after formal training has been completed. Three sources of information were synthesized in this report: (1) the current literature, (2) a meta-analytic review, and (3) an original empirical study investigating the relationship between IAT duration and skill-level retention in the SOF language training environment. Despite gaps in the empirical literature on this topic, results from the primary SOF study indicate IAT duration has an impact on SOF operators' initial proficiency; however, this impact lessens over time for both Category I/II and Category III/IV languages. Recommendations include further research on the relationship between initial training duration and skill-level retention, and a possible reduction in IAT length for Category I /II and Category III/IV languages based on the end-goal of IAT.

15. SUBJECT TERMS
Training, duration, retention, SOF, IAT

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Surface, Eric A.
U	U	U	UU (SAR)	42	**19b. TELEPHONE NUMBER** *(include area code)* 919-480-2751

Examine the Impact of Training Duration on Retention

SEPTEMBER 2012

APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION UNLIMITED
SPONSORED BY: SOFLO, USSOCOM
RESEARCH CONDUCTED BY: SWA CONSULTING INC.

EXECUTIVE SUMMARY

Issue and Overview

The ability to communicate in the target language is often critical to Special Operations Forces (SOF) mission success. According to the United States Special Operations Command (USSOCOM), "Language is a key component of SOF tasks" (USSOCOM M 350-8, p. 18). SOF operators suggest target language communication skills aid in training foreign nationals, persuading people to provide sensitive information, and maintaining control in hostile situations. Most importantly, SOF operators commonly describe instances when speaking the target language is helpful in building rapport with host nations. For example, one SOF operator noted, "Whenever you deploy anywhere, to work with anyone, and you don't have enough language capacity to build rapport, then you've got a problem" (*Inside AOR Use of Language* [Technical Report #2010011010]).

Despite the recognition of the importance of language capability in the SOF community, SOF language program administrators may face significant new challenges in developing and maintaining language-capable SOF personnel. In anticipation of Department of Defense (DoD) budgetary cuts and other possible budgetary constraints, this task may become even more challenging as program administrators are asked to do more with less in the near future (Steele, 2012). As a result of these budgetary cuts, SOF language program administrators may be called upon to shorten initial acquisition training (IAT).

To help SOF language program administrators obtain a better understanding of the long-term impact of shortening IAT on SOF operators' ability to retain their language capability in the future, the Special Operations Forces Language Office (SOFLO) sponsored this report to investigate the relationship between IAT duration and long-term language proficiency. The goal of this report is to provide SOF language program administrators with evidence-based (to the extent available) conclusions and recommendations pertaining to how IAT duration impacts SOF operators' ability to retain their language capability, specifically after formal training has been completed.

To accomplish this goal, this report synthesizes information across three sources to examine the relationship between training duration and long-term skill retention. These three sources of information are:

1. A thorough review of current literature available on the topic.
2. A meta-analytic review that aggregates findings from individual studies across different settings and populations to obtain a more accurate and robust estimate of the relationship between training duration and skill-level retention.
3. An original empirical study conducted in a SOF environment investigating the relationship between IAT duration and skill-level retention.

Conclusions

There is a significant gap in the current literature investigating the relationship between training duration and skill-level retention. This was most clearly evidenced in the difficulties encountered while conducting the meta-analytic portion of this report. More specifically, the severe lack of data investigating training

duration and skill retention in a longitudinal fashion, the presence and/or lack of control for significant confounding variables (e.g., length of retention interval), and failure to report adequate statistics to compute comparable change scores between time points (e.g., standard deviations, correlations between the measures at time 1 and time 2) negatively impacted our ability to conduct the meta-analysis. Further, most research on skill-level retention has focused on the retention interval, not on the relationship between training duration and skill-level retention. This gap in the literature was also apparent in the literature review, which was able to offer theoretical mechanisms that may be driving the relationship between training duration and long-term proficiency, but was able to offer little in terms of strong, empirically-derived information regarding the nature of this relationship or tests of these theoretical mechanisms.

By examining primary SOF data, however, we were able to find some evidence suggesting that IAT duration had a large impact on initial proficiency; however, over time, the differences in IAT duration had less and less of an impact on retention of language skills. This relationship was seen for both Category I/II and Category III/IV language difficulties. Thus, if the goal of training is to facilitate *short-term* retention of language skills, then a longer IAT duration may have a significant, positive impact towards achieving that goal. On the other hand, if *long-term* retention is the goal of training, then a longer IAT duration may be less important to long-term retention of language skills. More research, however, with larger and more robust datasets is of critical necessity to test this relationship. Additionally, more research is necessary to investigate the presence and impact of potential moderators of the relationship between training duration and skill-level retention.

Recommendations

Though much more research is needed, preliminary analyses indicate that differences in IAT duration had a lessening impact on long-term proficiency. More specifically, for Category I/II trainees, the difference in participatory listening proficiency scores between the 14-week and 18-week IAT duration groups was notable at the end of IAT but showed a downward trend over time and was minimal at Test 3. Similarly, for Category III/IV trainees, the difference in participatory listening proficiency scores between the 20-week and 24-week IAT duration groups was considerable at the end of IAT, but showed a downward trend over time and was minimal at Test 3. These preliminary results suggest it may be possible to reduce IAT duration to 14 weeks for Category I/II languages and 20 weeks for Category III/IV languages, if the goal of training is the *long-term* retention of language skills. These results, however, must be interpreted with caution because there were many potential moderators that were unable to be controlled for and the current study only compared two training durations. Future research investigating the recent change made at the U.S. Army John F. Kennedy Special Warfare Center and School (USAJFKSWCS) Basic Language Course (BLC) that modified IAT for Category I/II languages from 18 weeks to 24 weeks would offer a unique opportunity to investigate additional IAT durations and their impact on long-term skill retention.

This project was conducted by SWA Consulting Inc. under a subcontract with CACI-WGI, Inc. (Subcontract# B11-114482; Prime# H92222-10-D-0017/0007; Sub-CLIN 0003AB). For questions or more information about the SOFLO and this project, please contact Mr. Jack Donnelly (john.donnelly@socom.mil). For specific questions related to data collection or this report, please contact Dr. Eric A. Surface (esurface@swa-consulting.com) or Dr. Reanna Poncheri Harman (rpharman@swa-consulting.com).

TABLE OF CONTENTS

SECTION I: REPORT PURPOSE AND OVERVIEW

Examine the Impact of Training Duration on Retention Report Purpose

In the United States Special Operations Command (USSOCOM), Special Operations Forces (SOF) operators must acquire and maintain language capability to meet their language-related mission requirements and achieve mission success (*Inside AOR Use of Language* [Technical Report #2010011010], *Outside AOR Use of Language* [Technical Report #2010011011]). However, anticipated Department of Defense (DoD) budgetary cuts may result in corresponding resource reductions to SOF language training. As a result, SOF language program administrators may be called upon to shorten initial acquisition training (IAT) or identify ways to meet or exceed training objectives using fewer resources.

The primary purpose of this report is to provide USSOCOM with evidence-based (to the extent available) conclusions and recommendations regarding the long-term impact of shortening IAT on SOF operators' ability to retain their language capability in the future. To achieve this purpose, this report uses a three-pronged approach. First, it provides a review of the relevant theoretical and empirical literature available on this topic. Second, it presents and discusses the results of a meta-analysis that aggregates findings from individual studies across different settings and populations to obtain a more accurate and robust estimate of the relationship between training duration and skill-level retention. Third, it presents and discusses findings from an original empirical study investigating the relationship between training duration and skill-level retention, specifically in the SOF language training environment. Using elements from all three of these components, this report provides the Special Operations Forces Language Office (SOFLO) with guidance on how IAT duration impacts SOF operators' ability to retain their language capability, specifically after formal training has been completed.

Examine the Impact of Training Duration on Retention Report Overview

This report contains the following sections regarding the relationship between IAT duration and skill-level retention:

- Section II (p. 6) describes the rationale for this study.
- Section III (pp. 7-19) describes the theoretical and empirical literature on this topic.
- Section IV (pp. 20-28) presents the results of a meta-analysis that aggregates findings from individual studies across different studies and populations to obtain a more accurate and robust estimate of the relationship between training duration and skill-level retention.
- Section V (pp. 29-33) presents and discusses an empirical study investigating the relationship between training duration and skill-level retention, specifically in the SOF language training environment.
- Section VI (pp. 34-37) provides an overall synthesis and recommendations across Sections III-V (pp. 7-33).

SECTION II: STUDY RATIONALE

USSOCOM has a vision to have "SOF [who have]…basic through native language and culture capability" (USSOCOM M 350-8, p. 6). However, attaining language capability is difficult and the product of rigorous training and long-term efforts to retain, use, and improve language skills. For USSOCOM to achieve their vision of developing language-capable SOF operators, the SOF community must be able to accurately assess how both the length of formal IAT and the passage of time after formal IAT has concluded impact SOF operators' language capability and language-related mission readiness.

Despite the importance of language capability in the SOF community, SOF language program administrators face many challenges in developing language-capable SOF personnel. In the near future, this task may become even more challenging as program administrators are asked to do more with less in anticipation of DoD budgetary cuts and other possible budgetary constraints (Steele, 2012). The current report, in conjunction with a related report (*Examine Ways to Decrease Training Duration while Maintaining Training Objective* [Technical Report #2012010633]), investigates the role of time in the language acquisition and maintenance process to help inform strategy and policy that ensure SOF personnel are trained in the most efficient and effective manner. This report focuses on the question of how IAT duration impacts skill-level retention over time, while its associated report examines what factors can be leveraged to make training more efficient, while continuing to produce language-capable SOF personnel. These two reports investigate two different aspects of the same challenge: continuing to produce language-capable SOF personnel while working under tighter budgetary constraints. For USSOCOM to continue to successfully produce language-capable SOF personnel who maintain their language proficiency over time, it is critical for SOF leadership to understand how these budgetary realities may affect future language training efforts. This research strives to inform and provide that understanding by examining the following research questions:

- According to previous literature, what is the impact of training duration on skill-level retention?
- To what extent can previous research provide a more accurate understanding of the relationship between initial training duration and skill-level retention?
- What is the relationship between initial training duration and skill-level retention in the SOF language training environment?

SECTION III: LITERATURE REVIEW

The following section describes the theoretical and empirical literature relevant to the impact of training duration on skill-level retention. First, the basic learning processes will be discussed, followed by an examination of the possible impact of the passage of time on the basic learning process, and an overview of the contributions from both the general training literature and the language learning literature. Then, the discussion focuses on the key variables of interest to this report that are examined in the meta-analysis (Section IV, pp. 20-28) and original empirical study (Section V, pp. 29-33), and also identifies possible moderators of the relationship between the passage of time and skill-level retention.

The Basic Learning Process

The current literature on learning and skill acquisition was reviewed to gain insight into the language acquisition process. While numerous models of acquisition have been proposed (e.g., Encoding Specificity Theory, Cormier, 1987; The Dreyfus Model, Dreyfus & Dreyfus, 1986), most support a three-stage process of skill acquisition and development. This process is most often credited to Fitts and Posner (1967; Figure 1, p. 8). Examining learning as a linear process, Fitts and Posner's (1967) three-stage model of skill acquisition includes: (1) the cognitive phase, (2) the associative phase, and (3) the autonomous phase. In their model, it is assumed that with continued appropriate instruction and as time passes, skill levels will increase, and an individual will progress through the three phases. However, the time spent in one phase may not necessarily be equivalent to time spent in another. It is important to recognize skill acquisition as a continuous process, allowing for the existence of "overlap" between the stages of the learning process (Wrisberg, 2001).

The first, or cognitive, phase represents the beginning of the learning process. Within this stage, the learner develops a mental picture and a fuller understanding of the task at hand. During this stage of learning, errors are common and there is a reliance on external models of correctness. Because of this, it appears the learner is most influenced by instruction during this initial stage of learning (Fitts & Posner, 1967).

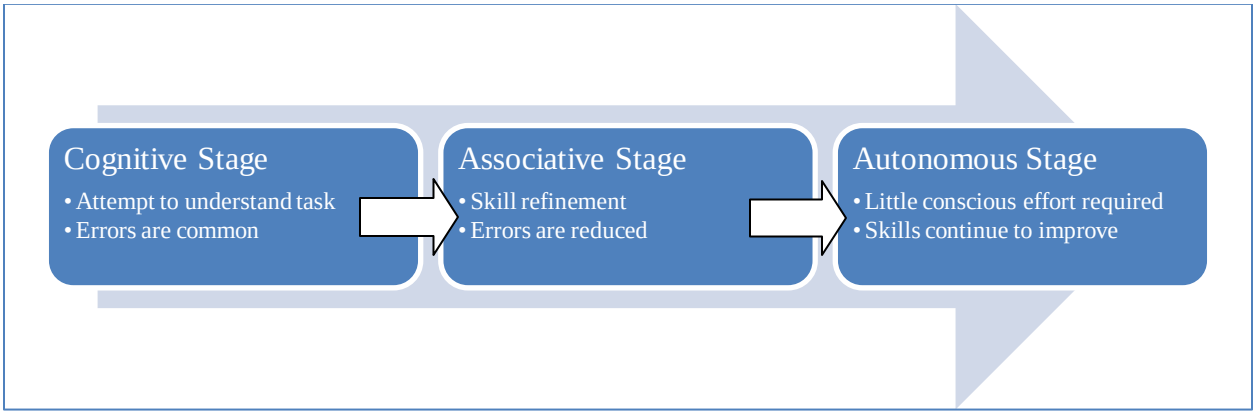
The second phase of Fitts and Posner's (1967) model is the associative phase, where the skills being learned are refined. As the learner enters this phase of the learning process, there is a gradual decrease in errors and the learner begins to rely more on an internal model, or reference, of correctness. It is during this phase when the learner is considered to have "successfully acquired the skill" (Delacruz, Chung, & Bewley, 2006, p. 5).

The final phase of the Fitts and Posner learning model is the autonomous, or automatic, phase where the learner is able to perform the skill(s) with little conscious effort (Fitts & Posner, 1967). In this phase, there is still skill improvement, although gradual, and it is plausible the learner may continue to improve indefinitely (Anderson, 1982).

Parallels can be seen between the phases of learning depicted by the Fitts and Posner model and levels of foreign language proficiency, as described by the Interagency Language Roundtable (ILR) scale. Figure 2 (p. 8) illustrates the ILR Levels and is presented in conjunction with Figure 1 (p. 8) to better demonstrate the parallels between Fitts and Posner's (1967) learning model and the ILR Levels.

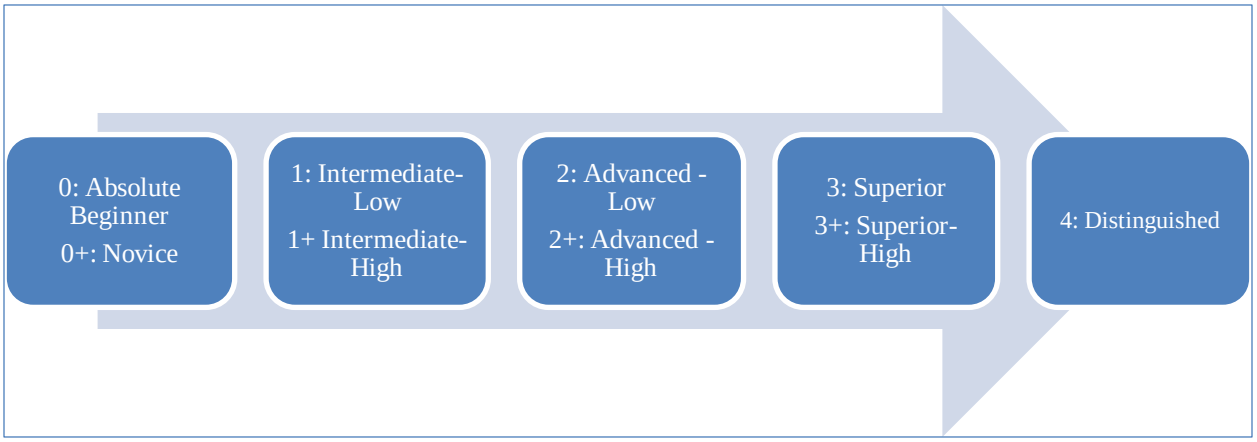
With some exceptions, most students start at a similar level of proficiency, typically 0 or 0+, at the beginning of IAT, similar to the cognitive phase of the Fitts and Posner model. As IAT progresses, over time and with adequately robust second language exposure and input, a learner will typically progress through the three stages and, it is assumed, progress to higher proficiency levels (e.g., 2+ to 4 on the ILR scale), although the pace of this progression will differ for each individual based on other factors and may not always be linear. These factors will be discussed in more detail below.

Figure 1. The Three Phases of the Learning Process (as proposed by Fitts and Posner, 1967)



Note. Skill acquisition is a continuous process; as such, there may be "overlap" between the stages of the learning process depicted in this figure (Wrisberg, 2001).

Figure 2. ILR Levels of Proficiency



Note. The progression through the ILR levels may not happen in evenly spaced increments, as suggested by this model.

The Impact of the Passage of Time on the Basic Learning Process

Previous research has not reached a consensus as to an optimal time necessary for initial foreign language acquisition to ensure long-term retention. However, research has heavily focused on the effect of the duration of the retention interval, defined as the time since formal training has ended and the time proficiency is re-tested (Bahrick, 1984), on level of language skill retained. Previous research has

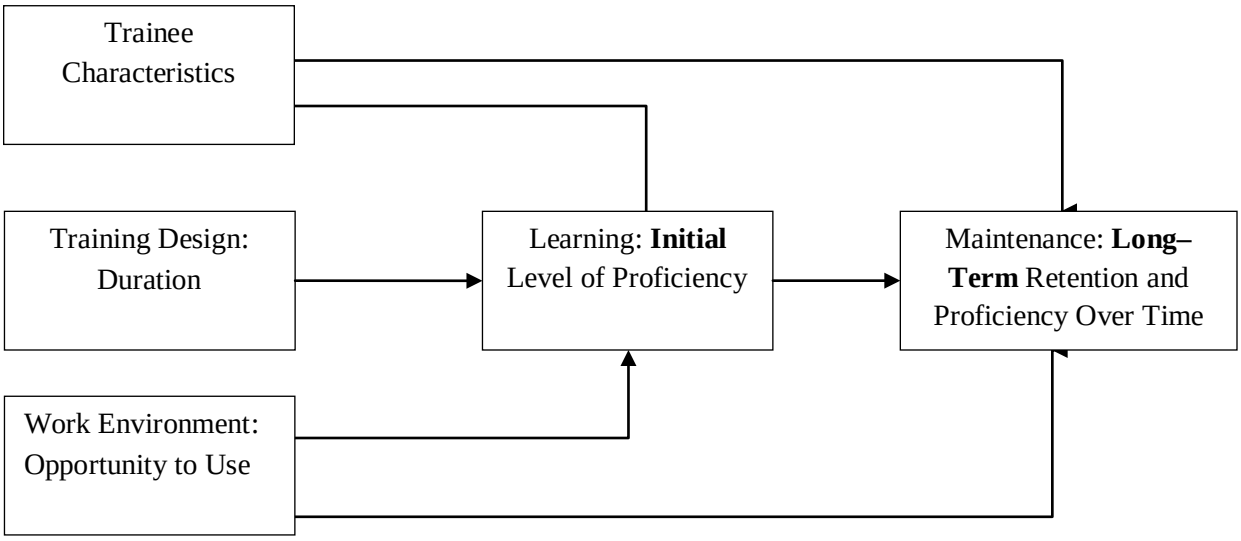
indicated there is, in general, a negative relationship between retention interval duration and language proficiency, such that proficiency levels tend to decrease as the duration of the retention interval increases (Bloomfield, Masters, Ross, O'Connell, & Gynther, 2012; Murtagh & van der Slik, 2004; Nagasawa, 1999; Reetz-Kurashige, 1999). In other words, the longer the time that passes after IAT ends, the lower proficiency scores tend to be on the re-test. This loss in skill proficiency has often been referred to as skill "attrition" (Bardovi-Harlig & Stringer, 2010), which is the opposite of skill maintenance or retention. Contributions to the discussion of what factors impact skill-level retention over time and how training duration might play a role, from both the broader training literature and the specific language learning literature, will be discussed.

Contributions from the Broader Training Literature

Investigating the retention of foreign language proficiency skills can also be viewed as an investigation of the degree to which students can successfully transfer skills learned in their training environments to their work environments and maintain those skills over time. Positive transfer of training is defined as the degree to which trainees effectively apply the knowledge, skills, and attitudes gained in a training context to the job (Newstrom, 1984; Wexley & Latham, 1981). The conditions of transfer include both the generalization of learned material to the job or other contexts and the maintenance of trained skills over a period of time on the job (Baldwin & Ford, 1988). This investigation is primarily concerned with the latter.

In their 1988 review of the transfer of training literature, Baldwin and Ford (1988) describe the general structure of the transfer process in terms of three groups of factors: (1) training-input factors, (2) training outcomes, and (3) conditions of transfer. The discussion below touches on each of the three factor groups, highlighting the importance of each in the SOF language training environment. Figure 3 (p. 10) draws from Baldwin and Ford's (1988) original model of the transfer process, and is adapted to the SOF language training context to illustrate the impact each group of factors may have on the long-term maintenance of language skills.

Figure 3. Transfer of Training Framework in the SOF Language Training Environment



Training Input- Duration and Trainee Characteristics

Baldwin & Ford (1998) identify training design as an important training-input factor. Examples of training design elements include training distribution (e.g., massed vs. distributed), presence of feedback, degree of overlearning, and training duration. The focus of this report is the aspect of training design concerned with training duration and how it may impact the transfer of language skills to the operational environment.

Trainee characteristics are another feature of training input identified in the Baldwin and Ford (1988) model that may impact skill retention over time. Although empirical research on the effects of ability, personality, and motivational effects were limited when this model was conceptualized (Baldwin & Ford, 1988), additional empirical research has been conducted on this topic. To that end, the effects of trainee characteristics, as presented in the language learning literature, are discussed in more detail later in this section.

Training Outcome: Initial Level of Proficiency

Training outcomes are defined as "…the amount of original learning that occurs during the training program and the retention of that material over a period of time on the job" (Baldwin & Ford, 1988, p. 64). The outcome of IAT is measured in terms of the level of target language proficiency achieved by each SOF operator. The initial level of proficiency achieved at the end of IAT will impact the skills available to a SOF operator for transfer to the operational environment. However, what may ultimately affect the impact of initial level of proficiency on the retention of skills may be the time between the end of training and opportunity for use.

Conditions of Transfer: Ability and Opportunity for Use

The work environment context refers to the work environment properties that function as constraints and opportunities to perform skills learned in the training environment on the job (Baldwin & Ford, 1988).

The "work" environment for SOF operators is much different than for other jobs and those discussed in the general training literature. For SOF operators, "on the job" may refer to their day-to-day duties and also to downrange mission-specific environments.

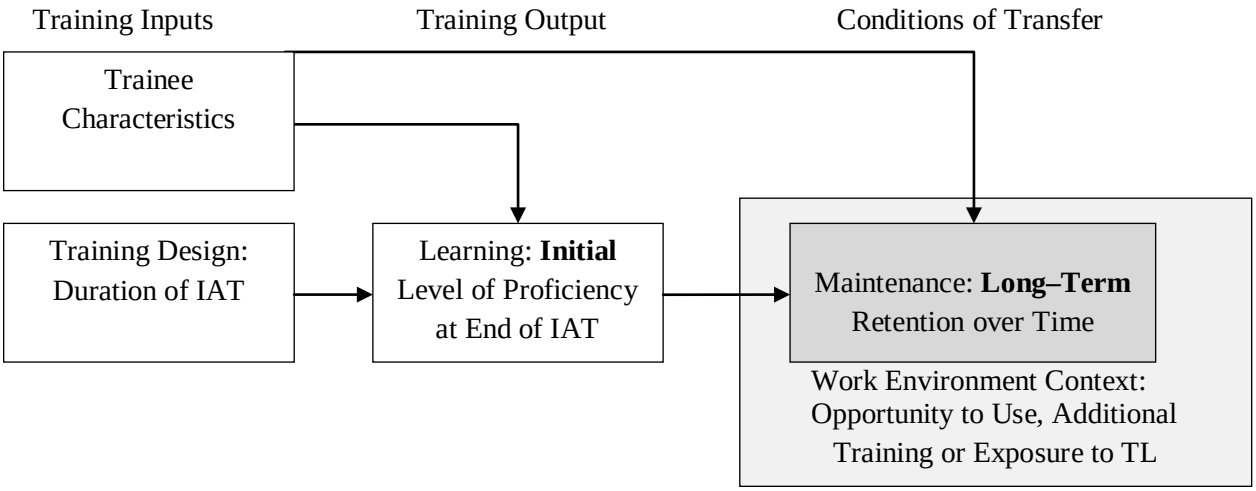
Figure 4 (p. 12) represents an adaptation of the Baldwin and Ford (1988) training transfer process; this adaptation highlights the longitudinal nature of the process and modifies it to better illustrate the transfer process in the SOF language training environment. In the adapted model:

- Training inputs are represented by trainee characteristics and the training design element of IAT duration.
- The initial level of proficiency at the conclusion of IAT represents the training output.
- The conditions of transfer are exemplified as the maintenance and long-term retention of language skills over time, which are operating within the larger context of the work environment.

This model suggests that trainee characteristics (i.e., training input) and training output have both direct and indirect effects on conditions of transfer, while IAT duration (i.e., another training input factor) has only an indirect effect on long-term retention through the initial level of proficiency reached at the end of IAT and the work environment context.

The larger context of the work environment is proposed to have the largest impact on long-term retention because both the training output (initial level of proficiency) and trainee characteristics operate within and through that context. Johns (2006) supported the important role of context and proposed that context can be conceived of as "a set of situational opportunities for, and countervailing constraints against . . . behavior" (Johns, 2006, p. 387). For example, a trainee with very high cognitive ability may have difficulty maintaining proficiency if the work environment context does not allow the opportunity or resources for the trainee to use his or her language skills after training has ended. For example, SOF operators often face constraints, such as limited or no resources and deployments to regions or countries where their languages are not spoken (USSOCOM M350-8, *Inside AOR Use of Language* [Technical Report #2010011010], *Outside AOR Use of Language* [Technical Report #2010011011]). Similarly, a SOF operator with a low initial proficiency level that has ample language learning resources available and is deployed multiples times to the target language region may be able to surpass his IAT proficiency scores. In this model, the impact of IAT duration has a direct influence on the training output (level of proficiency at end of IAT), but as time goes on, other elements, such as trainee characteristics and the larger context of the work environment, have increasingly stronger effects on long-term skill retention.

Figure 4. Transfer of Language Skills over Time



Note. Adapted from Baldwin & Ford (1988) Model of the Transfer of Training Process

Contributions from the Language Learning Literature

Previous literature specific to language learning has also identified several theories which offer potential mechanisms driving language skill retention over time, such as the simplification hypothesis (Bardovi-Harlig & Stringer, 2010), the regression hypothesis (as discussed by Vechter, Lapkin, & Argue, 1990; Bardovi-Harlig & Stringer, 2010), and the dormant language hypothesis (credited to Nelson, 1978). However, the theory of language retention that appears to be the most relevant to discuss as a possible mechanism driving the relationship between training duration and long-term retention is the critical threshold hypothesis (credited to Neisser, 1984), or the "best learned, last out" approach (Bardovi-Harlig & Stringer, 2010, p. 16).

According to this theory, if a certain threshold of use is achieved, skills may be less susceptible or vulnerable to attrition (Neisser, 1984). Simply put, "the more you know, the less you lose" (Hansen, 1999, p. 151). Research on language attrition has found that skill areas that are not "completely ingrained" attrite at a faster pace, despite some level of competence attained in these areas (Vechter et al., 1990, p. 297). But exactly what is this threshold? The American Council on the Teaching of Foreign Languages (ACTFL) and ILR scales are examples of these thresholds for varying language proficiency levels. Only two attempts (Hedgcock, 1991; Kennedy, 1932), however, have been made to operationalize the threshold concept empirically (Bardovi-Harlig & Stringer, 2010). However, the operationalization attempts of these authors were far too contingent upon their respective research questions to be generalizable to other studies, illustrating that the concept of the "threshold" may be sample- and context-dependent. The lack of research on the conceptualization of the "threshold" may be due to the difficulty of measuring a concept that may differ dramatically based on individual and environmental factors.

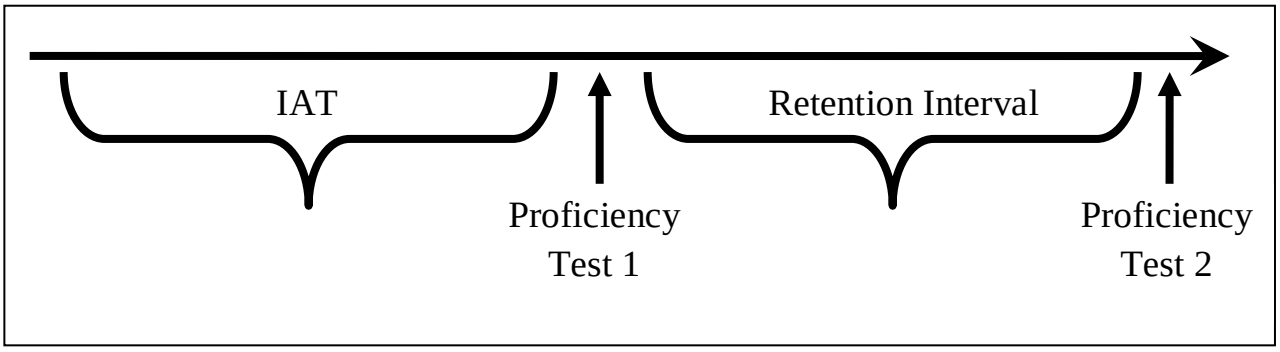
The threshold hypothesis illustrates the impact IAT duration may have on the retention of language skills over time. If IAT is not long enough in duration to allow SOF operators to reach the necessary "threshold" of target language skills, any skills acquired during that time period may be more vulnerable

to attrition. In other words, whether or not SOF operators reach the threshold necessary to help adequately fortify their skills against attrition depends, at least in part, on if they have been given enough time and training to reach that threshold.

Key Study Variables

The key variables of interest in this report include: (1) the length of IAT, (2) the length of the retention interval, and (3) the concept of language proficiency. Figure 5 (p. 13) depicts the primary variables in relation to the acquisition of language skills over time and further illustrates the context in which these variables will be discussed.

Figure 5. Key Variables in Language Skill Acquisition over Time



Length of IAT

The initial discussion and integration of Fitts and Posner's (1967) model and the literature on the threshold hypothesis of language learning indicate that the duration of the initial learning period is important, in terms of both its impact on initial proficiency and level of skill retained over time. The language learning literature has not identified a specific time period necessary to acquire proficiency, and therefore, has not identified a necessary duration of the initial acquisition period for retaining what is learned. However, estimates of the time to reach a certain level of proficiency have been specified. Throughout the language learning literature, the 1973 Foreign Service Institute (FSI) estimates appear to be the most widely-referenced guidelines and are based on experiences teaching students in FSI classes. The importance of time in the acquisition process has been highlighted numerous times in the language learning literature over the years (Diller, 1978; Jackson & Kaplan, 2001; Krashen, 1981), and overall, "…there is no substitute for simply spending time using the language" (Jackson & Kaplan, 2001, p. 77). If the level of proficiency achieved during IAT is one of the key variables related to the retention of language skills, then the length of initial training is important in allowing operators time to reach that "critical threshold" of proficiency to maintain their skills. Subsequently, from the training transfer perspective, the duration of IAT is a training input factor that may impact the level of skill available to SOF operators for transfer to the operational environment.

Length of Retention Interval

In the current study, the retention interval is operationalized as time from when formal training ends to when language proficiency skills are retested to determine if skill attrition has occurred. The length of the

retention interval examined in studies of language skill attrition varies. Previous studies have included intervals ranging from seconds and minutes, to the length of a three-month summer vacation (Smythe, Jutras, Bramwell, & Gardner, 1973), to periods of many years (Bahrick, 1984). In his study looking at language attrition, Bahrick (1984) examined the retention of foreign language skills in individuals who had completed college and/or high school Spanish courses from as few as one to as many as 50 years prior to his study. The authors concluded that with "…all other factors being equal (e.g., amount of foreign language use, motivation), ILR ratings for these individuals will tend to decrease as the amount of time between test administrations increases" (Bloomfield et al., 2012, p. 1354).

Language Proficiency Testing

Language proficiency tests serve as accurate and valid indicators of the nature, extent, and degree of effectiveness with which the examinee is able to communicate in the target language, as defined by the ILR language skill level descriptions (ILR, n.d.). Language proficiency tests can provide information about the level of skill developed during IAT and therefore, are used to determine whether the specific training objective has been achieved. Fulcher (2003) states that the purpose of proficiency testing is to "collect evidence in a systematic way that will support an inference about the construct" of interest based on test scores (p. 47). Proficiency test scores assess whether or not the intended skill (e.g., speaking proficiency) has been acquired. For SOF operators, proficiency testing at the end of training allows for inferences to be made about whether or not they have gained the target language speaking and listening skills necessary to achieve their mission tasks and objectives.

SOF operators are often tested using the two-skill Oral Proficiency Interview (OPI), which examines speaking and participatory listening proficiency. The Defense Language Proficiency Test (DLPT) is a non-participatory listening and reading proficiency test, with each skill modality assessed independently (Defense Language Institute Foreign Language Center [DLIFLC], 2010). However, there are other tests that have been utilized in the language learning literature. For example, Smythe et al. (1973) used locally developed French achievement tests to examine listening and reading comprehension of 9th grade students in Ontario. Previous research has also used self-assessment questionnaires (e.g., Clark & Jorden, 1984) to indirectly gauge language proficiency (Bardovi-Harlig & Stringer, 2010). The Test of English as a Foreign Language (TOEFL) administered by the Educational Testing Service(ETS) is used to assess English proficiency.

Key Moderators to be Examined

While a negative relationship between skill proficiency and retention interval duration has been largely supported by the literature, the level of proficiency loss over time may not occur in a completely linear fashion and other factors may moderate the relationship between the passage of time and target language skill attrition (Bahrick, 1984). To obtain a complete picture of how the passage of time affects target language proficiency, it is important to take these factors into consideration (Bloomfield et al., 2012). These factors may include learner activities during the retention interval (e.g., successful completion of Sustainment Enhancement Training [SET] activities, immersion opportunities), individual differences affecting the entire learning process (e.g., peak proficiency level achieved during IAT), and contextual or training design factors present in the operators' environment both during and after IAT (e.g., target language difficulty).

Level of Initial Proficiency

The initial level of target language proficiency an individual attains at the end of IAT may influence the long-term retention of language skills (Bahrick, 1984; Clark & Jorden, 1984), and it is one of the factors most discussed in the literature as affecting the retention of target language skills (Vechter et al., 1990). Previous studies have found the best predictor of retention to be the level of target language proficiency attained (Gardner, Lalonde, Moorcroft, & Evers, 1987). Similarly, according to de Bot and Clyne (1989), individuals reporting the most severe proficiency attrition initially had low proficiency (Hansen, 1999; Nagasawa, 1999; Reetz-Kurashige, 1999; Smythe et al., 1973).

However, there has been research that refutes the importance of initial target language proficiency level on retention. The results of the research conducted by Weltens ,van Els, and Schils (1989) suggest that attrition is independent of the training level of an individual, and their data points towards a fixed amount of attrition as opposed to a fixed proportion of the original knowledge, which is similar to the findings of Bahrick (1984).

Difficulty of Language Acquired

A common way to think about language difficulty is how different, or far away, the target language is from the learner's native language. In other words, the more difficult languages (those further away from English) will take longer to learn than those that are closer in nature to English. The four-category government classification system reflects the difficulty of a native English speaker to learn the target language (Silva & White, 1993; Surface, Dierdorff, & Donnelly, 2004). For example, the Category I languages Spanish and French share a similar alphabet with English and should be easier to learn for native English speakers than Category IV languages like Arabic or Japanese. This supports the idea that a more difficult language will require more time to learn and therefore more time to reach the "threshold." Therefore, the length of IAT necessary to reach the critical threshold may be different for more difficult to learn languages than it may be for less difficult to learn languages, and if the length of IAT is not adjusted based on the difficulty of the language learned, more difficult languages may attrite faster than easier to learn languages. If an individual already knows a language closely related to the target language, learning the target language will be easier (Diller, 1978).

Language Modalities Examined

Language modalities are the different ways people express themselves in a language. Some of the modalities that are commonly tested when determining proficiency are speaking, listening, reading, and writing. Specifically, at the end of the retention interval, individuals may be retested to determine the amount of initial skills that have been lost. However, the amount of attrition that is detected may be dependent on the type of test used, as well as the type of language modality retested (e.g., speaking versus listening proficiency).

The two-skill OPI is used to determine the speaking and participatory listening proficiency levels of SOF operators at the conclusion of IAT, and is also used to reexamine these proficiency levels after the retention interval. There are various other tests used in the literature to determine the proficiency of individuals after the retention interval. These tests may examine the retention of general language skills, such as reading or listening proficiency, or more specific grammatical language concepts.

Not all language modalities are affected by attrition in the same manner. For example Smythe et al. (1973) found a decrease in reading comprehension skills following summer break for 9th grade French students.

Individual Differences

Each individual has characteristics that are innate or constant over time, affecting their initial language learning success and their retention of language skills over time. In both the broader training literature and the language learning literature, a number of important individual differences have been extensively examined. For example, examinations in the broader training literature have focused on four general categories of individual differences: (1) capabilities (e.g., general mental ability), (2) demographics (e.g., age), (3) personality traits (e.g., goal orientation, general self-efficacy), and (4) interests and values (e.g., education; Gully & Chen, 2010; Kraiger, Ford, & Salas, 1993).

When determining the likelihood of success in language learning, individual characteristics such as age (e.g., Littlewood, 1984) and language learning experience (e.g., Jackson & Kaplan, 2001) must be examined as well. Some of these characteristics may be indirectly influenced through the training environment (e.g., motivation), but they more often than not have to be accommodated by other factors also affecting language acquisition (Littlewood, 1984), such as course and training environment factors. For more information about these and other individual differences in language learning success, refer to the *Examine Ways to Decrease Training Duration while Maintaining Training Objective* report (Technical Report #2012010633).

Conclusions

The literature review has attempted to highlight and explain the key variables and possible moderators of the relationship between the duration of IAT and the retention of language skills over time. To better understand the nature of this relationship, the report will now move to discuss the results of a meta-analysis, followed by an overview of an original empirical study conducted using SOF training effectiveness data.

Sections I-III References

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*, 369-406.

Bahrick, H. P. (1984). Semantic memory content in permastore: 50 years of memory for Spanish learned in school. *Journal of Experimental Psychology: General, 113*, 1-29.

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

Bardovi-Harlig, K., & Stringer, D. (2010). Variables in second language attrition: Advancing the state of the art. *Studies in Second Language Acquisition, 32*, 1-45.

Bloomfield, A. N., Masters, M. C., Ross, S. J., O'Connell, S. P., & Gynther, K. (2012). Change in foreign language skills over time. Retrieved August 15, 2012 from http://mindmodeling.org/cogsci2012/papers/0241/paper0241.pdf

Clark, J. L. D., & Jorden, E. H. (1984). A study of language attrition in former U.S. students of Japanese and implications for design of curriculum and teaching materials: Final Project Report. *Center for Applied Linguistics* (pp. 1-75). Washington, D.C.

Cormier, S. M. (1987). The structural process underlying transfer of training. In S. Cormier & J. Hagman (Eds.), *Transfer of learning: Contemporary research and applications* (pp. 151-181). San Diego, CA: Academic Press.

de Bot, K., & Clyne, M. G. (1989). Language reversion revisited. *Studies in Second Language Acquisition, 11*, 167-177.

Defense Language Institute Foreign Language Center. (2010). *Defense Language Proficiency Test 5 system familiarization guide for constructed response format.* Defense Language Institute Foreign Language Center. Retrieved September12, 2012 from http://www.dliflc.edu/archive/documents/Generic-Fam%20Guide-CRT-CBT.pdf

Delacruz, G. C., Chung, G.K ., & Bewley, W. L. (2006). Characterizing trainees in the cognitive phase using human performance knowledge mapping tool (HPKMT) and microgenetic analysis. CSE Technical Report 699: National Center for Research on Evaluation Standards, and Student Testing (CRESST).

Diller, K. C. (1978). *The language teaching controversy*. Rowley, Massachusetts: Newbury House.

Dreyfus, H. L., & Dreyfus, S. E. (1986). *Mind over machine: The power of human intuition and expertise in the age of the computer*. Oxford: Blackwell.

Fitts, P. M., & Posner, M. I. (1967). *Human performance*. Monterey, CA: Brooks-Cole.

Fulcher, G. (2003). *Testing second language speaking.* London: Longman/Pearson Education.

Gardner, R. C., Lalonde, R. N., Moorcroft, R., & Evers, F. T. (1987). Second language attrition: The role of motivation and use. *Journal of Language and Social Psychology, 6*, 29-47.

Gully, S., & Chen, G. (2010). Individual differences, attribute-treatment interactions, and training outcomes. In S. Kozlowski & E. Salas (Eds.), *Learning, training, and development in organizations* (pp. 3-64). New York, NY: Routledge.

Hansen, L. (1999). Not a total loss: The attrition of Japanese negation over three decades. In L. Hansen (Ed.), *Second language attrition in Japanese contexts* (pp. 142-153). Oxford: Oxford University Press.

Hedgcock, J. (1991). Foreign language retention and attrition: A study of regression models. *Foreign Language Annals, 2*, 43-55.

Interagency Language Roundtable. (n.d.). *Descriptions of proficiency levels*. Retrieved September 21, 2012 from http://www.govtilr.org/Skills/ILRscale1.htm.

Jackson, F. H., & Kaplan, M. A. (2001). Lessons learned from fifty years of theory and practice in government language teaching. In J. Alatis & A-H. Tan (Eds.), *Georgetown University round*

table on languages and linguistics 1999: Language in our time: Bilingual education and official English, ebonics and standard English, immigration and the Unz initiative (pp. 71-87). Washington, DC: Georgetown University Press.

Johns, G. (2006). The essential impact of context on organizational behavior. *Academy of Management Review, 31*, 386-408.

Kennedy, L. (1932). The retention of certain Latin syntactical principles by first and second year Latin students after various time intervals. *Journal of Educational Psychology, 23,* 132 – 146.

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

Kraiger, K., & Ford, J. K. (2010). Individual differences, attribute-treatment interactions, and training outcomes. In S. Kozlowski & E. Salas (Eds.), *Learning, training, and development in organizations* (pp. 3-64). New York, NY: Routledge.

Krashen, S. D. (1981). *Second language acquisition and second language learning.* New York, NY: Oxford University Press.

Littlewood, W. T. (1984). *Foreign and second language learning.* Great Britain: Cambridge University Press.

Murtagh, L., & van der Slik, F. (2004). Retention of Irish skills: A longitudinal study of a school-acquired second language. *International Journal of Bilingualism, 8*, 279-302.

Nagasawa, S. (1999). Learning and losing Japanese as a second language: A multiple case study of American university students. In L. Hansen (Ed.), *Second language attrition in Japanese contexts* (pp. 169-200). Oxford: Oxford University Press.

Neisser, U. (1984). Interpreting Harry Bahrick's discovery: What confers immunity against forgetting? *Journal of Experimental Psychology: General, 113*, 32 – 35.

Nelson, T. (1978). Detecting small amounts of information in memory: Savings for nonrecognized items. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 453-468.

Newstrom, J. W. (1984, August). *A role-taker time differentiated integration of transfer strategies.* Paper presented at the annual meeting of the American Psychological Association, Toronto, Ontario.

Reetz-Kurashige, A. (1999). Japanese returnees' retention of English-speaking skills: Changes in verb usage over time. In L. Hansen (Ed.), *Second language attrition in Japanese contexts* (pp. 21-58). Oxford: Oxford University Press.

Silva, J. M., & White, L. A. (1993). Relation of cognitive aptitudes to success in foreign language training. *Military Psychology, 5, 79-93.*

Smythe, P. C., Jutras, G. C., Bramwell, J. R., & Gardner, R. C. (1973). Second language retention over varying time intervals. *Modern Language Journal, 52*, 73-78.

Steele, D. (2012, April). Ups and downs in cutting back and moving forward. *Army Magazine, 62*(4), 34-40.

Surface, E. A., Dierdorff, E. C., & Donnelly, J. (2004, April). *Modeling second language proficiency change for US Special Operations personnel.* Poster presented at the 19th annual meeting of the Society of Industrial and Organizational Psychology, Chicago, IL.

SWA Consulting Inc. (2010, October). *Inside AOR Use of Language* (Technical Report #2010011010). Raleigh, NC: Author.

SWA Consulting Inc. (2010, October). *Outside AOR Use of Language* (Technical Report #2010011011). Raleigh, NC: Author.

SWA Consulting Inc. (2012, September). *Examine Ways to Decrease Training Duration while Maintaining Training Objective* (Technical Report #2012010633). Raleigh, NC: Author.

USSOCOM M 350-8. (2009). *Training: The Special Operations Forces language program.* MacDill Air Force Base, FL: Author.

Vechter, A., Lapkin, S., & Argue, V. (1990). Second language retention: A summary of the issues. *The Canadian Modern Language Review, 46*, 289-303.

Weltens, B., van Els, T. J. M., & Schils, E. (1989). The long-term retention of French by Dutch students. *Studies in Second Language Acquisition, 1*, 205-216.

Wexley, K. N., & Latham, G. P. (1981). *Developing and training human resources in organizations.* Glenview, IL: Scott Foresman.

Wrisberg, C. A. (2001). Level of performance skill: From beginners to experts. In R. Singer, H. Hausenblas, & C. Janelle (Eds), *Handbook of sport psychology* (pp. 3-19). New York, NY: John Wiley & Sons.

SECTION IV: META-ANALYTIC REVIEW

Purpose

Due to anticipated DoD budgetary cuts, SOFLO is interested in the potential impact of a reduction in the duration of IAT on skill-level retention. To help SOF language program administrators obtain a better understanding of the long-term impact of shortening IAT on SOF operators' ability to retain their language capability in the future, SOFLO sponsored this meta-analytic investigation. The primary goal of this investigation is to inform strategy and policy to ensure SOF personnel are trained in the most appropriate, efficient, and effective manner.

The purpose of the meta-analytic portion of this project is to quantitatively review the existing research literature and summarize how the retention of language skills changes as a function of initial training duration. This approach offers several significant advantages. First, meta-analysis combines the results of previous research to examine and statistically determine the average relationships across all included studies. Unlike individual research studies, meta-analysis provides a statistical summary of existing research to enhance the understanding of the relationships between variables (Hunter & Schmidt, 2004). Understanding the results of this prior research allows SOF leaders and language program administrators to make informed decisions regarding current SOF IAT practices. Prior research can also provide invaluable information about the current research questions, allowing detailed hypotheses to be formed. The current study examined over 2,000 references to foreign language retention to better understand the nature of the relationship between initial training duration and skill-level retention.

Secondly, a quantitative synthesis of this nature enables the researcher to not just note whether a significant relationship between initial training duration and skill-level retention exists, but also note the magnitude of the relationship (Schmidt, 1992). Understanding the magnitude of the relationship enables SOF leaders and language program administrators to make informed policy decisions related to SOF IAT and its recommended duration. Simply knowing whether or not initial training duration is related to skill-level retention is insufficient; a precise estimate of the nature of that relationship is of critical importance.

Another advantage to conducting this meta-analysis concerns the dynamic nature of the training duration variable. To detect, understand, or even begin to discuss its relationship with skill-level retention, there must be variability in initial training duration. However, typically, single studies will describe only one training program and will consequently have only one training length. The benefit of conducting a meta-analysis is researchers can more easily look at variables which are difficult or impossible to manipulate in a single study (Viswesvaran & Ones, 1995), such as multiple initial training lengths. In this case, meta-analytic techniques allow researchers to include a variety of studies examining the retention of language skills, even if single studies include only one training program with one training length.

Finally, conducting a meta-analysis of the existing research literature is a necessary step in providing evidence-based recommendations for practice (Briner & Rousseau, 2011). It is an important method to collect all available empirical evidence and can provide strong and robust conclusions.

Research Question

The fundamental research question is: *How does the length of a formal training program affect an individual's ability to retain the broad language skills learned during that training program?*

The three variables under investigation in this research question are: (1) the length of training, (2) an assessment of language skills at the end of that training, and (3) a further assessment of language skills at a later time point. These variables need to be operationalized in terms of effect sizes that can be the subject of the meta-analysis. Following practices from prior work on skill retention (e.g., Arthur, Bennett, Stanush, & McNelly, 1998), the *d*-value (i.e., standardized mean difference) between language skills after training and language skills measured later was meta-analyzed. The *d*-value is simply the difference in average skill proficiency between two time points, standardized to be on a common metric (i.e., a standard deviation metric). *D*-values have been calculated so that negative values indicate loss of proficiency. Next, the length of the training program was examined as a moderator of *d*-value. This moderator analysis is the focal analysis that addresses the research question.

Method

Search Strategy

Several keyword searches were conducted using multiple databases (e.g., Business Source Premier, Education Resources Information Center, Military and Government Collection, PsychINFO, and PsychArticles) using the following search terms: (training OR learning OR acquisition OR instruction OR teaching) AND (retention OR decay OR attrition OR maintenance OR sustainment) AND (language). The search was conducted in July of 2012. After removing duplicate entries, this search strategy resulted in 2,192 published and unpublished articles that were potentially relevant to the research question. The titles and abstracts of these articles were further hand-sorted for relevance; after eliminating irrelevant articles (e.g., cross-sectional studies, articles without primary data; see "Inclusion Strategy" for further details), 296 articles were collected for coding.

Additionally, a hand search of five journals (*Foreign Language Annals, Language Learning, Modern Language Journal, Second Language Research,* and *Studies in Second Language Acquisition*) was conducted to find published research that the initial search strategy did not capture. This yielded 16 additional articles.

To located unpublished studies providing relevant data, conference programs (American Association for Applied Linguistics [AAAL], ACTFL, Georgetown University Round Table on Languages and Linguistics [GURT], and Language Testing Research Colloquium [LTRC]) from the past 10 years were also searched, along with the websites of several organizations working in language testing, for relevant white papers or technical reports. After contacting authors to request papers, we received 11 additional studies.

Finally, email requests were sent out to several language-related listservs requesting unpublished work relevant to the research question; however, none of the responses provided data that could be included in the present meta-analysis.

Due to logistical and timing constraints, a small portion of studies identified in the searches could not be located in time for this report. These studies (e.g., theses/dissertations) represented only a small proportion of the identified articles (~5%). However, given the very low base rate of useable data, it is unlikely that substantial amounts of data are located in these sources.

Inclusion Criteria

To be included in this meta-analysis, studies had to meet the following criteria:

- The same knowledge and/or skills had to be measured consistently across at least two time points. Studies that collected data with some delay after training, but without a second assessment, were eliminated.
- No formal training on the focal knowledge/skills was reported between the two measurement time points.
- The study must use a sample of participants learning a foreign language (i.e., the criteria used should not measure proficiency in a sample's native language).
- The study had to include sufficient data for the calculation of a *d*-value (see below for details).
- The study must be written in English.

Additionally, studies were excluded from the meta-analysis based on the following criteria:

- Studies examining the retention of very narrow skills (e.g., a list of 20 vocabulary words) learned in very short time periods by experimental participants.
- The study was limited to individuals without learning or other types of disabilities.

Moderators

Time-related variables (i.e., initial training length and time between tests) were coded in weeks. Contact hours were unable to be coded due to lack of reporting in the primary studies. Language modality was coded into one of six categories: (1) Speaking, (2) Listening, (3) Reading, (4) Writing, (5) Mix, or (6) Other. The immersion context was coded into three categories: (1) immersed in L1, (2) some exposure to L2, or (3) immersed in L2. The language in which people were trained, as well as the native language, were recorded to estimate language learning difficulty. Unfortunately, due to a lack of primary studies, these moderators (e.g., language modality, language difficulty) could not be examined.

Further details on the methodology, including a more detailed account of the inclusion criteria, as well as details on the analysis and coding strategies, can be found in the Meta-analysis Technical Appendix (pp. 39-42).

Findings

Unfortunately, due to the inclusion criteria and both statistical and logistical challenges, only four studies with four independent samples could be included in the meta-analysis. Despite our best efforts to include what we could, sufficient data simply are not reported in the research literature that would allow us to include more studies in this meta-analysis. To supplement the lack of studies, the results from the original empirical SOF study (Section V, pp. 29-33), as well as a review of key findings from articles not included in the meta-analysis, were included.

The four studies included in this meta-analysis all share a particular set of characteristics. All were published by Robert Gardner and colleagues, and all studied Canadian high school students learning French. One study was for six weeks of IAT, one study was for 39 weeks of IAT (this study comprised 220 students), and the two remaining studies had training lengths in the 195-364-week range. These four studies used non-military samples and criteria. Some care must be taken in the interpretation of these findings, as differences between these samples and the military samples from the SOF primary data analyses can introduce confounds. In Table 1 (p. 23), note that the results for the final two IAT length categories (i.e., 29-39 weeks and 195-364 weeks) are driven entirely by the samples of Canadian high school students learning French.

Table 1 (p. 23) contains the results of the meta-analysis, and Figure 6 (p. 24) visually presents these results. There is no clear trend in the data, but it does appear that longer training duration does result in less language skill decay. The *d*-values for training lengths of 29 weeks and above are very small (*d* = -.10 for 29-39 weeks, and *d* = -.12 for the longer duration). This indicates that there is only a nominal decrease in skill with very long training lengths. If this were to be interpreted on an IQ metric, for example, this would be a decrease of only 1.5 points.

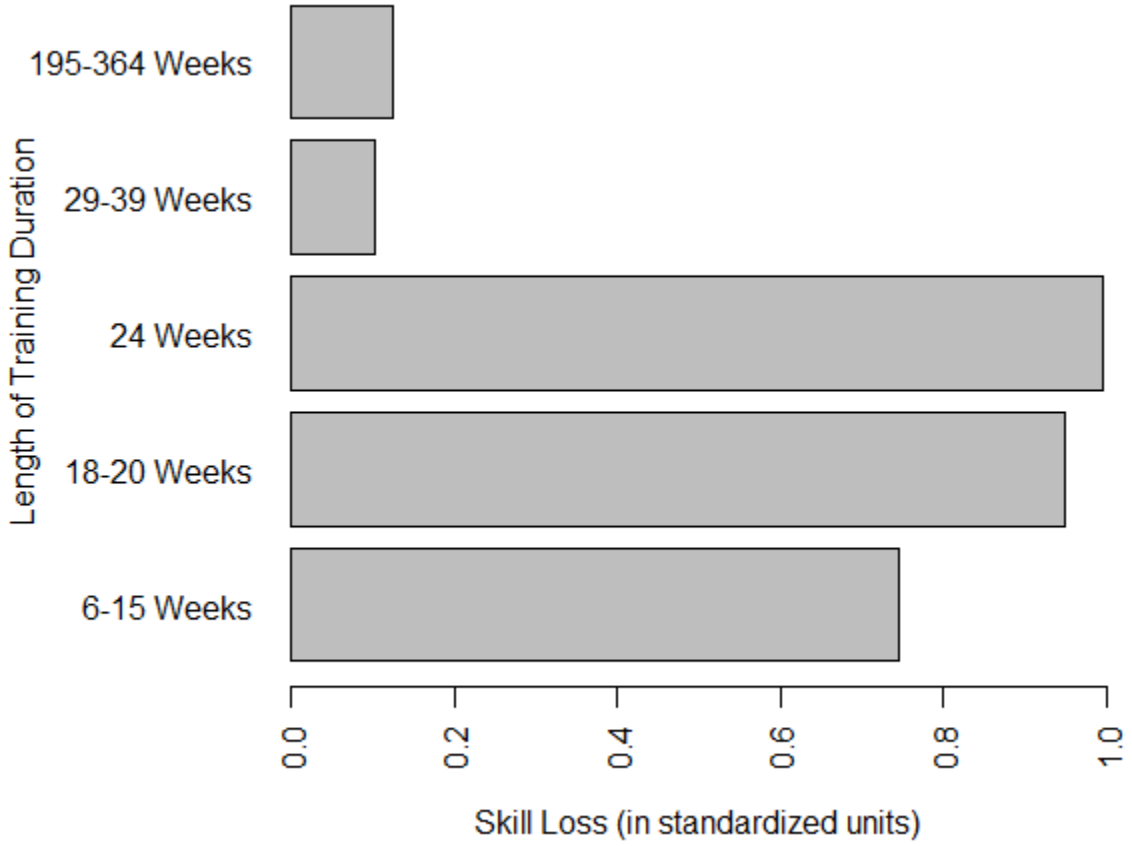
On the other hand, shorter training lengths appear to lead to significantly more decay. For the three shorter training lengths, *d*-values range from -.75 to -1.00. These are large and indicate a rather substantial amount of language skill decay. Again, translating this to an IQ metric, these *d*-values are the equivalent of an 11-15-point drop in skill. This is rather substantial and noticeable.

Table 1. Meta-analytic Results

Initial Training Length (weeks)	***K***	***N***	**Mean *D***	***SD* obs.**	***SD* res.**	**90% CV**
6-15	4	599	-0.744	0.236	0.048	[-0.82, -0.66]
18-20	7	1,719	-0.949	0.367	0.127	[-1.16, -0.74]
24	6	1,329	-0.996	0.592	0.343	[-1.56, -0.43]
29-39	3	228	-0.102	0.009	0.000	[-0.10, -0.10]
195-364	2	194	-0.124	0.126	0.004	[-0.13,- 0.12]

Note: K = number of studies. *N* = total sample size. Mean *D* = sample-size-weighted average *d*-value. *SD* obs. = observed standard deviation of *d*-values. *SD* res. = residual standard deviation once sampling error has been accounted for. 90% CV = lower and upper bounds of the 90% credibility interval.

Figure 6. Length of Training Duration and Skill Loss over Time



Initially, there seems to be a counterintuitive trend in the data, with more skill decay as training length increases. Unfortunately, there is a confound with language difficulty, such that all studies training for 24 weeks were Category III/IV languages, and all but one of the studies for shorter training periods were Category I/II languages. An alternative explanation of that trend is that is it more difficult to retain the language skills from more difficult languages. Due to scarcity of studies contributing to the meta-analysis, we were unable to examine potentially relevant moderators (e.g., language difficulty).

Weighted Regression Analyses

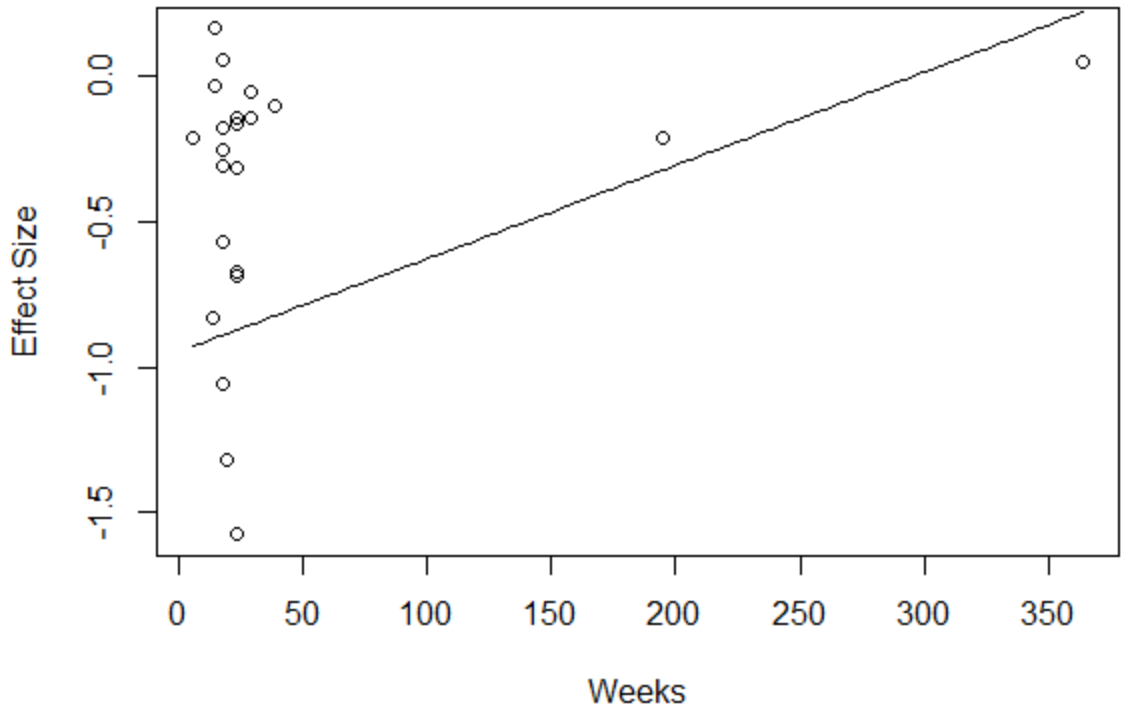
To further examine how the length of training is related to retention, a weighted regression analysis was conducted on the studies included in the meta-analysis, with the *d*-value as the dependent variable and training duration as the independent variable. Data points were weighted by the sample size, as *d*-values based on larger sample sizes are more stable.[1]

Although the results were not statistically significant due to the small number of studies that could be included, there are some interesting findings. The data seem to indicate that a longer training duration leads to less skill decay over time (R^2 = .115, p = .123).

[1] The relationship between training duration and skill decay over time can be described by the following equation: (*d-value*) = -0.95 + .0032(*# of weeks*)

Figure 7 (p. 25) shows the plot of the effect sizes vs. duration of IAT (in weeks), with the regression line from above added. The confidence in the regression results must be tempered, though, as they are heavily influenced by two outlier studies with very long training durations (i.e., several years in duration). Upon removing these studies, the relationship disappears. However, the results of those studies are included, as they provide interesting findings not currently captured in the rest of the data.

Figure 7. Scatterplot of Effect Sizes vs. Training Duration



Supplemental Review of Studies

Several studies were not included in the meta-analysis due to a lack of sufficient statistics available to compute a *d*-value. This section is a review of the most relevant of these studies to supplement the findings from the meta-analysis. For a more detailed review of language attrition research in general, several excellent reviews exist (e.g., de Bot & Weltens, 1995; Oxford, 1982; Pan & Gleason, 1986; Weltens, 1987), including a special issue on the topic from the journal *Studies in Second Language Acquisition* (Weltens & Cohen, 1989).

The second study of Smythe et al. (1973) found that French students waiting to continue language learning until the Fall semester the following year had a significant drop in skills; however, the magnitude of this drop was approximately equal for students with 1-3 years of high school French. Counter to expectations, students' language skills actually improved when they continued language study in the Fall

semester. This effect was found no matter the initial acquisition duration; however, the authors point out that these findings may be the result of individual differences in motivation to learn a second language.

Cohen (1974) studied 14 children after an 18-week Spanish immersion program. After a summer break, he noted there was a significant decline in speaking skills, with nouns being the most subject to decay. This would suggest that 18 weeks of training is not sufficient to maintain language skills over time, even over a fairly brief retention interval of a few months.

Bahrick (1984) examined a large number of adults who had studied Spanish at some point in their lives (usually high school or college) and had since ceased Spanish instruction. His results indicated that the longer it had been since training, the more skills decayed, as expected. This effect was not moderated, however, by the length of training in Spanish. Similar levels of language skill decay were observed for all training levels.

These supplemental reviews suggest several additional findings. After sufficient training, language skills can be maintained over shorter time frames (i.e., 2-3 months), even with no formal instruction. However, language skills will begin to decay over longer intervals. With shorter training lengths (e.g., 18 weeks), skills begin to decay fairly quickly, with basic vocabulary (i.e., nouns) being one of the first language skills to decay. Somewhat counter to the results of the meta-analysis, Bahrick (1984) found that while longer training durations led to higher skills no matter how long it had been since training ended, the rate of decay was the same for all training lengths. This issue merits further study, however.

Summary

This meta-analysis, in isolation, provides little in the way of actionable recommendations. The data in the current research literature that met our rigorous inclusion criteria are too sparse to responsibly draw any strong conclusions. By including a large amount of primary data with the data from the meta-analysis, however, we were able to draw some somewhat tenuous conclusions.

There is a slight general positive trend, in that longer training durations tend to lead to better skill retention and less skill decay. However, this relationship is heavily influenced by two studies which find very little skill day after long periods of training (up to near fluency). Once these studies are removed, this relationship effectively disappears. With so few studies available for analysis, exploratory analyses cannot be supported.

Interpretability is further clouded by the inability of this study to examine any potentially relevant moderators. Several key moderators (e.g., second language, second language use, language difficulty), though coded for, could not be analyzed due to insufficient data. In a sufficiently large meta-analysis, these issues tend to average out so that main effects can be interpreted. However, in a small-scale meta-analysis like this, these variables may introduce confounds in the data, thus weakening our ability to draw strong conclusions.

Language skill maintenance is a complex and multifaceted issue with many moving parts that have the ability to impact skill retention over time. Many individual (e.g., second language usage, motivation, language aptitude) and contextual variables (e.g., deployment location, setting, support to maintain language skills) have strong and substantial effects on language skill maintenance, along with training duration. These effects interact in complex ways; looking at a single variable in isolation (i.e., training

duration) does not give a complete or clear picture of how language skills may change over time as a function of that variable. This is evidenced in the difficulties encountered while conducting this meta-analysis, more specifically, a severe lack of data investigating training duration and skill retention in a longitudinal fashion, the presence and/or lack of control for significant confounding variables (e.g. length of retention interval), and failure to report adequate statistics to compute comparable change scores between time points (e.g. standard deviations, correlations between the measures at time 1 and time 2).

Unfortunately, due to these challenges and the fact that the tenuous conclusions that were able to be drawn were conflicting and fraught with interpretability issues, relatively little can be learned from this specific meta-analysis. These results are indicative of a significant gap in the language learning literature. In an effort to begin the formidable task of closing this gap, an original study that leverages historical and current SOF language training data was undertaken. This study has the ability to control for some of these confounding variables as well as using longitudinal data to provide a much clearer picture of how training duration impacts language skill retention over time. However, further study of this topic is warranted and necessary to tease apart the specific effects of training duration on the retention of language skills over time.

Section IV References

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

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

Bahrick, H. P. (1984). Fifty years of second language attrition: Implications for programmatic research. *Modern Language Journal, 68*, 105-118.

Briner, R. B., & Rousseau, D. M. (2011). Evidence-based I-O psychology: Not there yet. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 4*(1), 3-22.

Cohen, A. D. (1974). The Culver City Spanish immersion program: How does summer recess affect Spanish speaking ability? *Language Learning, 24*, 55-68.

de Bot, K., & Weltens, B. (1995). Foreign language attrition. *Annual Review of Applied Linguistics, 15*, 151-164.

*Gardner, R. C., Lalonde, R. N., Moorcroft, R., & Evers, F. T. (1987). Second language attrition: The role of motivation and use. *Journal of Language and Social Psychology, 6*, 29-47.

*Gardner, R. C., & Lysynchuk, L. M. (1990). The role of aptitude, attitudes, motivation, and language use on second language acquisition and retention. *Canadian Journal of Behavioural Science, 22*, 254-270.

*Gardner, R. C., Moorcroft, R., & Metforda, J. (1989). Second language learning in an immersion programme: Factors influencing acquisition and retention. *Journal of Language and Social Psychology, 8*, 287-305.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd Ed.). Thousand Oaks, CA: Sage.

Oxford, R. L. (1982). Research on language loss: A review with implications for foreign language teaching. *Modern Language Journal, 66,* 160-169.

Pan, B. A., & Gleason, J. B. (1986). The study of language loss: Models and hypotheses for an emerging discipline. *Applied Psycholinguistics, 7,* 193-206.

Schmidt, F. L. (1992). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologist, 47,* 1173-1181.

*Smythe, P. C., Jutras, G. C., Bramwell, J. R., & Gardner, R. C. (1973). Second language retention over varying intervals. *Modern Language Journal, 57,* 400-405.

Viswesvaran, C., & Ones, D. S. (1995). Theory testing: Combining psychometric meta-analysis and structural equations modeling. *Personnel Psychology, 48,* 865-885.

Weltens, B. (1987). The attrition of foreign-language skills: A literature review. *Applied Linguistics, 8,* 22-38.

Weltens, B., & Cohen, A. D. (1989). Language attrition research: An introduction. *Studies in Second Language Acquisition, 11,* 127-133.

SECTION V: SPECIAL OPERATIONS FORCES STUDY

Purpose

SOFLO is seeking ways to save money and allocate resources more efficiently. The reduction of IAT constitutes one potential way of cutting costs. However, if reducing training were to adversely influence language capability, this could potentially outweigh financial benefits that could be gained. Thus, it is necessary to first study the effects of initial training duration on proficiency over time before arriving at a decision about whether to reduce IAT length. To that end, the purpose of this primary data investigation was to explore the relationship between IAT length and language skill retention. In contrast to typical investigations, which focus on factors that influence language proficiency immediately following IAT, this research was concerned with the influence of initial training duration on the retention of language skills, long after the conclusion of initial acquisition training.

In contrast to the meta-analytic studies (Section IV, pp. 34-37), all samples used in this study are from SOF IAT events. This carries the primary advantage in that characteristics that differ between populations (e.g., age, aptitude, and gender) that could potentially influence results are held relatively constant.

Research Question

The research question for this investigation was: *How is the length of initial acquisition training related to the retention of language skills over time?*

Main Findings

- For Category I/II languages, SOF personnel who received 18 weeks of initial language training had higher DLPT Listening (DLPT-L) scores than personnel who received 14 weeks of training. This difference occurred for three time points: (1) post-IAT, (2) retention test 1, and (3) retention test 2. However, this difference in DLPT-L scores diminished over time, with very little difference in DLPT-L scores measured at the retention test 3 time point.

- For Category III/IV languages, SOF personnel who received 24 weeks of initial language training had higher DLPT-L scores than personnel who received 20 weeks of training. This difference occurred for all four time points: (1) post-IAT, (2) retention test 1, (2) retention test 2, and (3) retention test 3. Much like the findings for Category I/II languages, this difference diminished over time.

- Regardless of the difficulty of the language being studied, SOF personnel who received four additional weeks of initial language training had higher DLPT-L scores than SOF personnel who received four less weeks of initial language training. However, this proficiency gap diminished over time. Indeed, for the third retention test, which occurred an average of 196 weeks (almost four years) after initial language training, DLPT-L scores were roughly equivalent for SOF personnel who received 14 or 18 weeks of initial language training, as well as for SOF personnel who received 20 or 24 weeks of initial language training. Thus, these preliminary results suggest it may be possible to reduce the length of IAT SOF personnel who are assigned to Category I/II and Category III/IV languages receive by as many as four weeks without having a significant

impact on their long-term retention of these language skills (i.e., retention measured roughly four years later). However, due limitations of this study, additional investigation is necessary.

Recommendations

- For Category I and II languages, the IAT length could be 14 weeks, if the goal of training is to achieve satisfactory long-term retention of language skills (i.e., skill retention roughly four years after IAT).[2]

- For Category III and IV languages, the IAT length could be 20 weeks, if the goal of training is to achieve satisfactory long-term retention of language skills (i.e., skill retention roughly four years after IAT). [3]

- Additional research is needed to further examine the relationship between initial training duration and language proficiency over time in SOF language training environments. Recently, IAT for Category I and II languages at SWCS BLC was extended from 18 weeks to 24 weeks. As more data are collected from SOF operators enrolled in these training events, the relationship between training duration and language proficiency over time can be reexamined to shed additional light on this topic and more specifically, make comparisons in language proficiency scores between SWCS BLC Category I/II trainees who received 24 weeks of language training and SWCS BLC Category I/II trainees who received 18 weeks of language training.

- A consistent finding from the language attrition literature is that skill maintenance over time is very strongly influenced by second language usage after the conclusion of formal IAT. It is recommended that IAT should focus on strategies to maintain usage and practice when not actively engaged in field operations. Continued use of the target language, whether informally or as part of SET, is vital in the maintenance of those skills, even if this use is not part of a formal training program. Likewise, it is also recommend that language immersion opportunities or language events during Robin Sage be provided immediately after IAT.

Method

From 2004-2012, training effectiveness data were collected from Special Operations Forces (SOF) personnel from the Air Force Special Operations Command (AFSOC) and the U.S. Army John F. Kennedy Special Warfare Center and School (USAJFKSWCS) Basic Language Course (BLC). [4] Retention data from other SOF components and units were insufficient and therefore, were not included in this study.

[2] If the end goal of training is to maximize language proficiency in the short-term, then a longer IAT duration may have a significant, positive impact towards achieving that goal.
[3] If the end goal of training is to maximize language proficiency in the short-term, then a longer IAT duration may have a significant, positive impact towards achieving that goal.
[4] Between 2004 and 2012, there were notable changes in the structure of language training. For example, from 2007 to 2009, SWCS language training occurred in two short "blocks," with breaks in between, followed by one long "blitz" block. This differed from more recent SWCS language training (as well as SWCS language training that predated 2007), which occurred as one long, continuous training session. Such structural differences were often confounded with the total duration of training and thus could affect the interpretation of results in this study.

Language proficiency was measured by the listening portion of the DLPT (DLPT-L) in ILR units. Proficiency was measured four times; the first measurement time point occurred immediately after SOF IAT, and the remaining three measurement time points occurred after the initial post-IAT test. Thus, these three final measurement time points constituted SOF personnel's retention scores. Retention scores were only included in analyses if they occurred at least six months or longer following the post-IAT test and were in the same language as the post-IAT test. Due to insufficient retention data, this study did not use OPI scores as measures of language proficiency.

The influence of initial training duration on the retention of language proficiency was examined for two groups of SOF personnel. The first group studied easier to learn languages (e.g., Category I or II languages), and the second group studied more difficult to learn languages during their IAT (e.g., Category III or IV languages). For each group, data were only available for two IAT durations: (1) 14 and 18 weeks for SOF personnel who studied Category I/II languages, and (2) 20 and 24 weeks for SOF personnel who studied Category III/IV languages (Table 2, p. 31).

Table 2. Length of Retention Intervals for Category I/II and III/IV Languages

Retention Intervals	Average # of Weeks	Minimum # of Weeks	Maximum # of Weeks
Post-IAT Test to Retention Test 1			
Category I/II	84	26	329
Category III/IV	80	26	342
Post-IAT Test to Retention Test 2			
Category I/II	142	44	348
Category III/IV	143	40	324
Post-IAT Test to Retention Test 3			
Category I/II	196	81	361
Category III/IV	196	77	349

Results[5]

Category I and II Languages

Results for SOF personnel who studied Category I/II languages are shown in Figure 8 (p. 32) for two initial training durations: (1) 14 weeks, (2) and 18 weeks. In terms of their post-IAT, first retention, and second retention DLPT-L scores, SOF personnel who received 18 weeks of training in Category I/II languages had higher DLPT-L scores than SOF personnel who received 14 weeks of training. However, the difference in proficiency test scores for the SOF personnel who received 18 weeks of training in Category I/II languages and personnel who received 14 weeks of training diminished over time. Specifically, the difference in proficiency test scores was greater than 0.50 ILR units post-IAT, less than 0.50 ILR units for retention tests 1 and 2, and almost 0 ILR units for retention test 3. Sample sizes for these analyses are presented in Table 3 (p. 32).

[5] To conduct these analyses, training effectiveness data from different SOF components and units were combined. Thus, any differences in proficiency may be at least partly attributed to factors that varied between SOF components (e.g., instructors, structure and design of the training).

Figure 8. Differences in DLPT-L Scores Post-IAT and over Time for Category I/II Languages

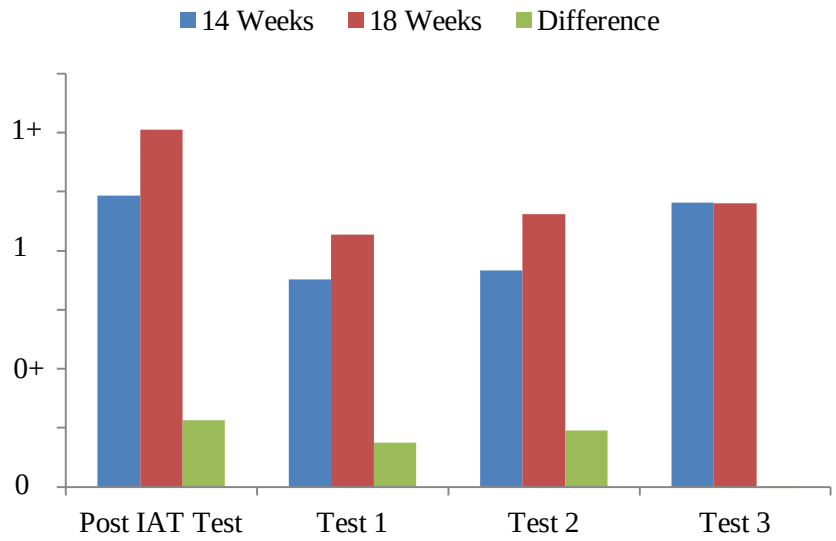


Table 3. Sample Size for each Training Duration for Category I/II DLPT-L Scores

Training Duration	Post-IAT Test	Test 1	Test 2	Test 3
14 Weeks	624	528	358	183
18 Weeks	1519	1060	716	412

Category III and IV Languages

Results for SOF personnel who studied Category III/IV languages are shown in Figure 9 (p. 33) for two initial training durations: (1) 20 weeks, (2) and 24 weeks. Consistent with results for SOF personnel who received training in Category I/II languages, personnel who received four additional weeks of training (i.e., 24 weeks) in Category III/IV languages had higher DLPT-L proficiency scores than personnel who received four less weeks (i.e., 20 weeks) of training in Category III/IV languages. However, the difference in proficiency test scores for SOF personnel who received 24 weeks of training in Category III/IV languages and personnel who received 20 weeks of training diminished over time. For retention test 3, DLPT-L scores were roughly equivalent between 24-week and 20-week trainees. Sample sizes for these analyses are presented in Table 4 (p. 33).

Figure 9. Differences in DLPT-L Scores Post-IAT and over Time for Category III/IV Languages

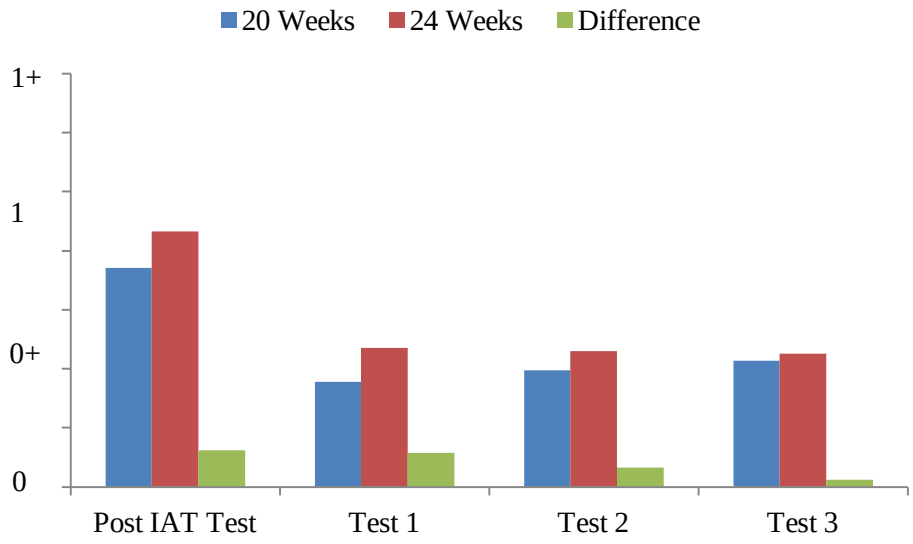


Table 4. Sample Size for each Training Duration for Category III/IV DLPT-L Scores

Training Duration	Post-IAT Test	Test 1	Test 2	Test 3
20 Weeks	630	504	319	130
24 Weeks	1355	950	628	318

SECTION IV: OVERALL SYNTHESIS & RECOMMENDATIONS

The goal of this report was to investigate the relationship between IAT duration and skill-level retention. To accomplish this goal, this report gathered and synthesized information across three sources. This type of systematic, multi-modal approach has been advanced as a necessary step in providing evidence-based recommendations for practice (Briner & Rousseau, 2011). The purpose of this section is to synthesize the primary conclusions and recommendations from the following sources of evidence:

1. A thorough review of current literature available on the topic.
2. A meta-analytic review that aggregated findings from individual studies across different settings and populations to obtain a more accurate and robust estimate of the relationship between training duration and skill-level retention.
3. An original empirical study conducted in a SOF environment study that investigated the relationship between initial training duration and skill-level retention.

Main Conclusion 1

The lack of empirical data found in both the literature review and the meta-analysis revealed a significant gap in the current literature in terms of empirical investigations of the relationship between initial training duration and skill-level retention.

- *Recommendation:* Simply stated, more research needs to be conducted on the relationship between initial training duration and skill-level retention. More specifically, there were some stark deficiencies in the current literature identified by this study that weakened the ability to draw strong recommendations based on the current literature and meta-analytic data. Below are some specific recommendations for future research that would help strengthen the current literature on this topic.
    - Conduct more rigorous, empirical research and report adequate statistics to compute comparable change scores between time points (e.g., standard deviations, correlations between the measures at time 1 and time 2).
    - Conduct longitudinal research to enable the investigation of proficiency at multiple points over time. This is not possible with only a cross-sectional design.
    - Use broader measures of language proficiency (e.g., not measuring only short-term memorization of 20 vocabulary words or a set of specific idioms).
    - Utilize research designs able to control for significant confounding variables (e.g., length of retention interval, language difficulty, individual differences).
    - When possible, use larger sample sizes to create more robust datasets and facilitate researchers' ability to draw stronger conclusions.

Main Conclusion 2

As suggested by the adaptation of the Baldwin and Ford (1988) transfer of training model, results from the original empirical study conducted in the SOF environment indicated that initial training duration had a large impact on initial proficiency; however, over time, the differences in IAT duration had less and less of an impact on retention of language skills. More specifically, for Category I/II trainees, the difference in participatory listening proficiency scores between the 14-week and 18-week IAT duration groups was notable at the end of IAT but showed a downward trend over time and was minimal at Test 3 (Figure 10, p. 35). Similarly, for Category III/IV trainees, the difference in participatory listening proficiency scores between the 20-week and 24-week IAT duration groups was considerable at the end of IAT, but showed a downward trend over time and was minimal at Test 3 (Figure 11, p. 36).

Figure 10. Differences in Language Proficiency Scores over Time for Category I/II Languages

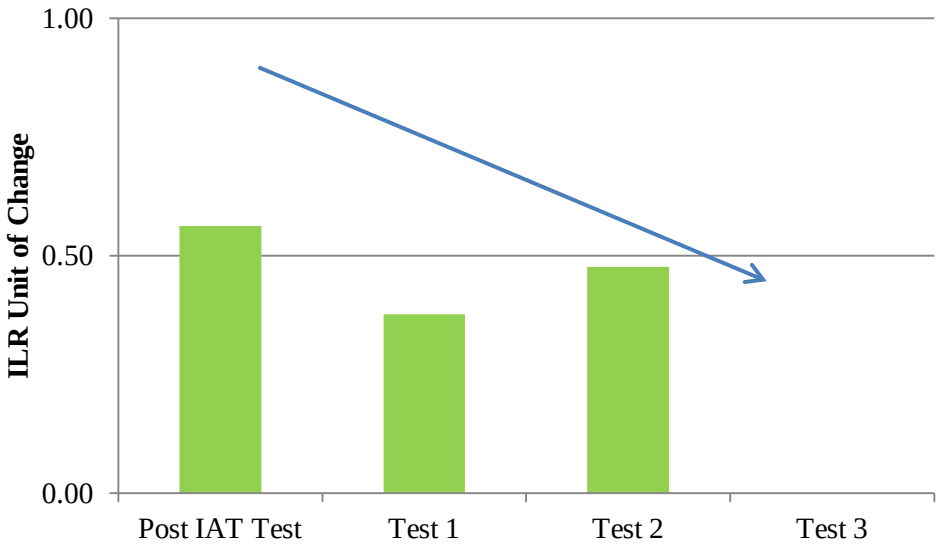
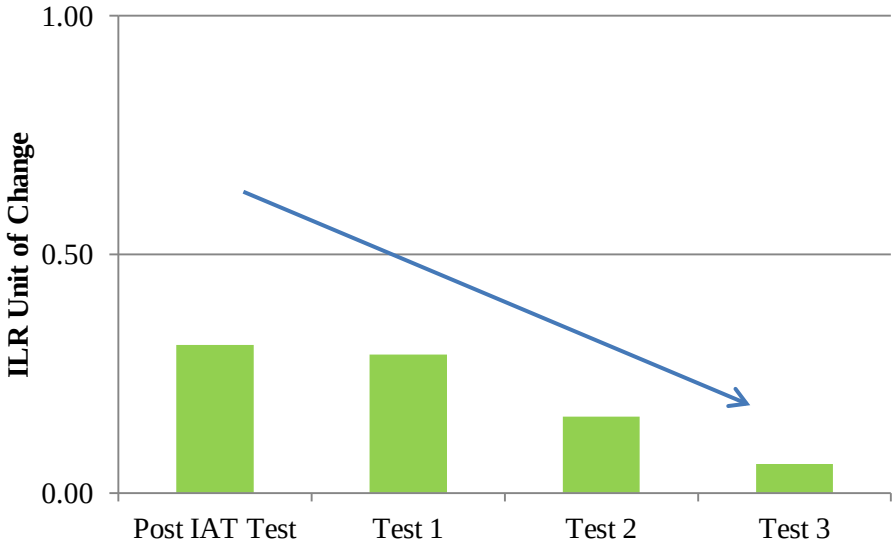


Figure 11. Differences in Language Proficiency Scores over Time for Category III/IV Languages



The practical implications of these findings will be different based on the end-goal of the training. If the goal of training is to facilitate *short-term* retention of language skills, then a longer IAT duration may have a significant, positive impact towards achieving that goal. On the other hand, if *long-term* retention is the goal of training (such as it was in this investigation), then a longer IAT duration may be less necessary for long-term retention of language skills.

- *Recommendation:* Preliminary results suggest it may be possible to reduce IAT for Category I/II languages to 14 weeks without having a significant, negative impact on long-term retention. Similarly, results indicated that it may be possible to reduce IAT for Category III/IV languages to 20 weeks without having a significant, negative impact on long-term retention. However, it should be noted there will a negative impact on the short-term retention of language skills if the IAT durations for Category I/II languages and Category III/IV languages are shortened to 14 weeks and 20 weeks, respectively.

These results, however, must be interpreted with caution because there were many potential moderators that were unable to be controlled for and only two training durations were compared for each group (i.e., Category I/II languages and Category III/IV languages). More research is needed to further investigate these findings before offering firmer recommendations.

Directions for Future Study

In addition to incorporating the above-referenced recommendations in an effort to improve the lack of statistical rigor and study design issues that permeate the current literature on this topic, a lack of variability in initial training duration was also identified as a substantial problem; however, the recent change made at SWCS BLC that modified the Category I/II language IAT from 18 weeks to 24 weeks offers a unique opportunity to further investigate the relationship between initial training duration and skill-level retention.

As more data are collected from SOF operators enrolled in these training events, the relationship between initial training duration and skill-level retention can be reexamined to shed additional light on this topic and more specifically, make comparisons in language proficiency scores between SWCS BLC Category I/II trainees who received 24 weeks of language training and SWCS BLC Category I/II trainees who received 18 weeks of language training, over both the short- and long-term.

Section VI References

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

Briner, R. B., & Rousseau, D. M. (2011). Evidence-based I-O psychology: Not there yet. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 4*(1), 3-22.

ABOUT SWA CONSULTING INC.

SWA Consulting Inc. (formerly Surface, Ward, and Associates) provides analytics and evidence-based solutions for clients using the principles and methods of industrial/organizational (I/O) psychology. Since 1997, SWA has advised and assisted corporate, non-profit and governmental clients on:

- Training and development
- Performance measurement and management
- Organizational effectiveness
- Test development and validation
- Program/training evaluation
- Work/job analysis
- Needs assessment
- Selection system design
- Study and analysis related to human capital issues
- Metric development and data collection
- Advanced data analysis

One specific practice area is analytics, research, and consulting on foreign language and culture in work contexts. In this area, SWA has conducted numerous projects, including language assessment validation and psychometric research; evaluations of language training, training tools, and job aids; language and culture focused needs assessments and job analysis; and advanced analysis of language research data.

Based in Raleigh, NC, and led by Drs. Eric A. Surface and Stephen J. Ward, SWA now employs close to twenty I/O professionals at the masters and PhD levels. SWA professionals are committed to providing clients the best data and analysis upon which to make evidence-based decisions. Taking a scientist-practitioner perspective, SWA professionals conduct model-based, evidence-driven research and consulting to provide the best answers and solutions to enhance our clients' mission and business objectives. SWA has competencies in measurement, data collection, analytics, data modeling, systematic reviews, validation, and evaluation.

For more information about SWA, our projects, and our capabilities, please visit our website (www.swa-consulting.com) or contact Dr. Eric A. Surface (esurface@swa-consulting.com) or Dr. Stephen J. Ward (sward@swa-consulting.com).

The following SWA Consulting Inc. team members contributed to this report (listed in alphabetical order):

Mr. Matthew J. Borneman
Mr. David D. Fried
Dr. Amanda Gissel
Mrs. Cristina Lambert
Dr. Jennifer Lindberg McGinnis
Dr. Eric A. Surface

APPENDIX: META-ANALYSIS TECHNICAL APPENDIX

Detailed Inclusion Strategy

The goal was to identify primary studies that examined the retention of foreign language knowledge and skills over a specified time frame, with the retention interval occurring after an identifiable period of initial instruction. To this end, and to be in alignment with the research question, several criteria were established to ensure consistency in coding to provide maximally interpretable results.

Due to the longitudinal nature of the research question, the first inclusion criterion concerned the nature of data collection. The same knowledge and/or skills had to be measured consistently across at least two time points. Studies that collected data with some delay after training, but without a second assessment, were eliminated. Additionally, the measure of knowledge/skill had to be consistent across time points; either identical measures or equated measures needed to be used so that scores would be on the same scale. Studies using two different tests with no mention of a procedure to ensure comparability of test characteristics (e.g., test difficulty) were not included.

A second criterion was that the study needed to focus on participants from a non-clinical population. This was to align the meta-analysis with the goals of the research question, which is concerned with the acquisition and retention of a second language in a non-clinical population of learners (e.g., SOF operators). As such, studies with clinical (e.g., autistic learners, individuals with learning disabilities) samples were eliminated.

Another key criterion for inclusion was that the study had to focus on participants learning a new language that was not their native language. Studies where people learned English as a foreign or second language were included, as were studies where American Sign Language was learned as a second language. However, a substantial number of studies examined memorizations of word lists and/or vocabulary in the leaner's native language; these studies were eliminated from inclusion in the meta-analysis a priori.

The fourth inclusion criterion focused on the nature of the research question. Because the research question is focused on the retention of foreign language skills after training has ended, we only included studies which had no formal training on the focal knowledge/skills between the two measurement time points. For example, studies which collected data at the end of one semester then again at the beginning of the next were included, assuming no additional training occurred during the interval (e.g., summer school). Studies in which data were collected at one time point during training, had the training continue, and then collected data at a subsequent point during training were excluded; these studies focused more on learning than on retention. Additionally, studies were included if there were some informal environment that might enhance the ability to retain language skills. For example, if people were immersed in a region where the target language was spoken, but were not actively taking part in a formal training program, that study was included in the analysis.

A fifth criterion was to only include studies that assessed language skills at a broad level. Again, this is to be consistent with the research question, which is concerned with the acquisition and retention of the

broad set of language skills that are developed in an intensive training program. As such, we did not include studies examining the retention of very narrow skills learned in very short time periods. These studies are typically focused on memory, rather than the integration of a broad set of skills and, as such, are a qualitatively different research question. Studies eliminated from this inclusion criterion typically taught a very narrow criterion (e.g., 40 vocabulary words) and for a very short period of time (e.g., under two hours).

Finally, to be included in the meta-analysis, the study had to include sufficient data for the calculation of a *d*-value (see below for details). Unfortunately, this proved to be a rather strict (though very necessary) inclusion criterion. Several clearly relevant studies (e.g., Study 2 of Smythe et al., 1973) were unable to be included because they did not report sufficient statistics to calculate a *d*-value. To mitigate the deleterious effects of being unable to include these studies in the meta-analysis, a qualitative review of the key studies that were unable to be included is provided.

Analytic Methodology

The statistic being accumulated for this meta-analysis was the within-subjects/repeated measures *d*-value. The formula for this statistic is:

$$d = \frac{\mu_2 - \mu_1}{\sigma_1}$$

where μ_1 and μ_2 are the means for time 1 and time 2, respectively, and σ_1 is the standard deviation for time 1. This is the most appropriate formula for the within-subjects/repeated measures *d*-value, as it has a known sampling distribution (Becker, 1988; Morris, 2000). Following other meta-analyses on skill retention (e.g., Arthur et al., 1998), the *d*-values were coded so that a negative *d*-value represented a loss in skill from time 1 to time 2 and positive *d*-values represent a gain in skill across time points. When converting to a *d*-value from test statistics, we paid careful attention to the direction of the effect to ensure consistency with this coding.

In the coding process, coders were careful to capture the correlation between the measures at time 1 and time 2, as this correlation affects the magnitude of the *d*-value in the conversion from test statistics and plays a large role in the calculation of the standard error (Dunlap, Cortina, Vaslow, & Burke, 1996). When this statistic was not available (and it was unavailable for the vast majority of the studies), the mean of all other values reported was used.

None of the articles coded provided a *d*-value. To compensate for this, the *d*-value was calculated directly from the reported means and standard deviations. When t-tests were provided, these were converted to *d*-values (Dunlap et al., 1996). When an article reported multiple effect sizes on the same sample, a composite *d*-value was computed to maintain independence of data (Hunter & Schmidt, 2004). In these cases, the formula from Sackett and Ellingson (1997) was used.

We used a Hunter-Schmidt (2004) style approach to this meta-analysis. It was modified slightly to use the more accurate estimates of the distribution of the within-subjects/repeated measures *d*-value (Morris, 2000; Viechtbauer, 2007). Following best practices, this is a random-effects model for meta-analysis (Schmidt, Oh, & Hayes, 2009). Additionally, it was decided to use a bare-bones method of meta-analysis,

not correcting for range restriction or unreliability. Because of the nature of the data contributing to the meta-analysis, range restriction is unlikely to be present to any significant degree. While it is recognized that unreliability is almost certainly present (e.g., Schmidt & Hunter, 1996), it has been explicitly chosen not to be corrected for because the research question is on the effects of measured proficiency, not theoretical effects.

Coding

Coding was carried out by two novice and two experienced meta-analysts. Before training began, one of the experienced coders began sorting a subset of the initial pool of articles into articles with codable data and those without. Once a small set of articles were identified that had codable data, five articles were chosen with which to conduct coder training to represent a diverse set of issues in the meta-analysis.

Coder training took place over several days. The first day was spent familiarizing everyone with the research question, coding sheet, and discussing the variables to be coded. Additionally, the experienced coders shared personal strategies for coding. Over the next few days, all four people independently coded the same five articles identified for training, then came together and discussed coding in a final meeting. Disagreements were few and were resolved with discussion. These discussions were used to refine the inclusion strategy and to develop a shared frame-of-reference for coding. Once this frame of reference had been sufficiently developed, the remaining articles were distributed among coders; each article was coded by one person. Any subsequent questions regarding coding were handled through collaborative discussion amongst two or more coders. All data contributing to the meta-analysis was subsequently verified by a second person before being analyzed.

References

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

Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology, 41*, 257-278.

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1*, 170-177.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd Ed.). Thousand Oaks, CA: Sage.

Morris, S. B. (2000). Distribution of the standardized mean change effect size for meta-analysis on repeated measures. *British Journal of Mathematical and Statistical Psychology, 53*, 17-29.

Sackett, P. R., & Ellingson, J. E. (1997). The effects of forming multi-predictor composites on group differences and adverse impact. *Personnel Psychology, 50*, 707-721.

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

Schmidt, F. L., Oh, I.-S., & Hayes, T. L. (2009). Fixed versus random-effects models in meta-analysis: Model properties and an empirical comparison of differences in results. *British Journal of Mathematical and Statistical Psychology, 62*, 97-128.

Smythe, P. C., Jutras, G. C., Bramwell, J. R., & Gardner, R. C. (1973). Second language retention over varying intervals. *Modern Language Journal, 57*, 400-405.

Viechtbauer, W. (2007). Approximate confidence intervals for standardized effect sizes in the two-independent and two-dependent samples design. *Journal of Educational and Behavioral Statistics, 32*, 39-60.