

Foreign Language Training Transfer: Individual and Contextual Predictors of Skill Maintenance and Generalization

J. Kemp Ellington

Illinois Institute of Technology

Eric A. Surface

SWA Consulting Inc., Raleigh, North Carolina

Brian D. Blume

University of Michigan, Flint

Mark A. Wilson

North Carolina State University

Foreign language proficiency is a critical skill in which many U.S. military personnel receive extensive training. However, very little research has examined the factors associated with the successful transfer of this training. This study therefore investigates the impact of individual and contextual variables on two different types of foreign language skill transfer measures in a military context. Archival data were analyzed from 133 U.S. Army Special Forces (Green Berets) teams, including 919 Soldiers who had completed job-required foreign language training. Results indicate that initial skill acquisition had a positive impact on both the maintenance and generalization of language skills. The posttraining time interval between training and transfer measurement was negatively associated with skill maintenance, suggesting significant skill decay over time. The team context also accounted for significant variability in skill transfer, and the team mean skill level moderated the relationship between individual initial skill and subsequent generalization to job performance.

Keywords: training transfer, foreign language training, skill maintenance, skill generalization, team context

J. Kemp Ellington, Department of Psychology, Lewis College of Human Sciences, Illinois Institute of Technology; Eric A. Surface, SWA Consulting Inc., Raleigh, North Carolina; Brian D. Blume, School of Management, University of Michigan, Flint; Mark A. Wilson, Department of Psychology, North Carolina State University.

An earlier version of this archival study was presented at the annual meeting of the American Psychological Association (APA) in San Diego, CA, and is based in part on a study presented to the Defense Language Action Panel sponsored by the Special Operations Forces Language Office (SOFLO), U.S. Special Operations Command (USSOCOM). This archival study drew upon data from a study of U.S. Army Special Forces field performance conducted by North Carolina State University (NCSU) and the U.S. Army Research Institute (ARI) for the U.S. Army Special Forces Command (USASFC) and U.S. Army John F. Kennedy Special Warfare Center (USAJFKSWCS) and School in FY 2001 and from the Defense Manpower Data Center (DMDC). SWA Consulting Inc. (SWA) conducted the archival study for SOFLO, USSOCOM. Throughout the article, the terms study and research are used interchangeably. The usage of these words is not meant to imply that

the investigation drew upon any particular source of funding. Furthermore, the views expressed in this article are those of the authors and do not necessarily reflect the positions or policies of NCSU, the ARI, the USASFC, the USAJFKSWCS, the DMDC, the SOFLO, the USSOCOM, the U.S. government, or SWA. Mark A. Wilson was the principal investigator for the field performance study. Eric A. Surface was principal investigator for the SOFLO sponsored study and has a financial interest in SWA, and he was also a Consortium Research Fellow for ARI during the original field performance study. We thank all the team members involved in the original field performance study, especially Jat Thompson and Mike Sanders. We thank the SWA team members involved in the archival study. We also thank all the sponsors of the original and archival studies. Finally, we thank Armando Estrada and the anonymous reviewers.

Correspondence concerning this article should be addressed to J. Kemp Ellington, Department of Psychology, Lewis College of Human Sciences, Illinois Institute of Technology, 3105 South Dearborn Street, Chicago, IL 60616. E-mail: jellingt@iit.edu

If learning is the proximal goal of job training, then the *transfer* of learning to performance in the work context is the more distal goal. Moreover, this extent to which training leads to meaningful changes in job performance is regarded as the “paramount concern of training efforts” (Baldwin, Ford, & Blume, 2009, p. 41). Because of globalization and geopolitical events, the foreign language communication needs of business, government, military, and educational organizations have become more salient in the U.S. Therefore, foreign language learning and job performance are becoming increasingly important to many organizations. For example, in certain parts of the U.S., companies in various industries are training their managers and customer service representatives to speak Spanish because an increasing percentage of employees and customers have Spanish as their primary language. With respect to the training and job performance of U.S. military personnel, foreign language proficiency is a skill domain of particular importance in the current global landscape (U.S. Department of Defense, 2010). In recent years there has been a growing recognition of the importance of language capabilities, with numerous U.S. government reports highlighting the need for and value of foreign language proficiency. For example, the 2005 Defense Language Transformation Roadmap (U.S. Department of Defense, 2005), the 2010 Quadrennial Defense Review (QDR), and a 2011 report from the Government Accountability Office (GAO) to congressional committees all stressed the importance of developing foreign language capabilities for both strategic and tactical purposes. In order to operate efficiently and effectively, today’s military personnel must be able to communicate and interact with local populations from numerous regions all over the world while conducting a variety of missions, ranging from training foreign militaries and conducting joint operations to humanitarian and aid operations. Having sufficient language capabilities increases both the likelihood of mission success and the degree of success, while decreasing risk to U.S. forces through greater situational awareness and more personal relationships with local personnel. Recent assessments within the U.S. Department of Defense (DOD) suggest language skills “are as important as critical weapon systems” (U.S. DOD, 2005, p. 3). Given the U.S. military’s substan-

tial investment in foreign language training (GAO, 2011), it is essential to examine the factors associated with the successful transfer of language skills to jobs and missions.

The concept of training transfer has been defined as including “both the *generalization* of trained skills to the job and the *maintenance* or long-term retention of trained knowledge and skills [italics added]” (Baldwin et al., 2009, p. 51). Accordingly, researchers and practitioners have used a variety of measures to operationalize transfer (Barnett & Ceci, 2002; Blume, Ford, Baldwin, & Huang, 2010), including those aimed at assessing knowledge/skill persistence over time as well as the application to job performance in the workplace. Historically, the U.S. military has taken a “maintenance” approach to evaluating foreign language training transfer, conducting ongoing posttraining proficiency assessments to measure the persistence of language skills over time (Dierdorff & Surface, 2008). Such an approach assesses trainees’ posttraining language “capabilities,” in that proficiency as measured using a standardized assessment provides an inference of the language skill level an individual is capable of applying to the performance of their job. However, an alternative way of approaching the measurement of language transfer would be to directly assess the extent to which trainees are effective in generalizing their foreign language skills to the performance of their job/mission tasks, such as communication activities during joint operations or training with foreign military personnel. This strategy focuses on the extent to which trained skills are applied in the work context, and assumes that performance can be measured directly by some method (e.g., supervisory ratings).

Though each of the above approaches assess important dimensions of transfer (Baldwin et al., 2009), the manner in which transfer is defined and operationalized is an important aspect of training evaluation, since learned/maintained skills do not necessarily generalize to performance in the job context (Blume et al., 2010). Although existing research has examined language training transfer as skill maintenance (e.g., Dierdorff & Surface, 2008), research on military populations to date has yet to examine transfer with regard to language-related job performance. To provide a more comprehensive understanding of language

transfer, it is critical to evaluate factors associated with *both* the maintenance of skills over time and the generalization of these skills to performance in the field. In addition, practical concerns such as the posttraining time interval between training and the measurement of transfer may impact skill maintenance and generalization, and thus may have important implications for language training effectiveness research and practice. For example, although proficiency testing for maintenance is currently required annually in the military system, many factors—such as deployments, other operational considerations, and test availability—can impact the time between assessments. Finally, it has been suggested that more robust explanations for behavioral processes can be derived by examining influences from different levels (e.g., individuals and teams; Hackman, 2003). Accordingly, numerous scholars have called for a multilevel approach to training effectiveness research (Aguinis & Kraiger, 2009; Kozlowski, Brown, Weissbein, Cannon-Bowers, & Salas, 2000), which involves analyzing training data using hierarchical modeling to capitalize on the nested nature of most learning data (e.g., learners within sections; sections within courses; courses within programs; etc.). Although previous studies have demonstrated that team contexts can impact training outcomes (Kozlowski et al., 2000; Smith-Jentsch, Salas, & Brannick, 2001), the degree to which team characteristics affect language transfer is currently undetermined.

With these issues in mind, the goals of this study were threefold. First, we provide a comprehensive examination of foreign language training transfer in a military setting, including measures of both language skill maintenance and the generalization of language-related training to the job. As part of this investigation, we examine the extent to which posttraining language acquisition predicts both dimensions of transfer, while controlling for trainee cognitive ability as well as the difficulty of the trained language. Second, addressing calls for further longitudinal studies of training transfer (Baldwin et al., 2009), we assess the impact of the duration of time posttraining on both skill maintenance and generalization measures. Previous studies on time intervals or maintenance curves have predominantly included transfer measures within six months to one year posttraining, and

have rarely focused on cognitive skills such as language proficiency (Baldwin et al., 2009; Blume et al., 2010; Taylor, Russ-Eft, & Chan, 2005). Importantly, this study incorporates a wide range of time intervals, allowing for a unique analysis of language proficiency transfer over time. Third, we take a multilevel analytic approach in order to investigate research questions surrounding team context influences on transfer, including the potential moderating role of team mean language skill on individual-level transfer relationships.

In order to address our goals, we examined these relationships using archival data obtained from a sample of U.S. Army Special Operations Forces (ARSOF) personnel, specifically U.S. Army Special Forces (SF; Green Berets). Because SF Soldiers have a foreign language requirement, they are required to train and maintain their foreign language skills. Additionally, though SF Soldiers use language to support their core missions, they are not linguists, for whom language capability is their core mission. Thus, SF foreign language transfer research should more readily generalize to a variety of military and civilian jobs, where language enhances the core job tasks and activities, as opposed to research with linguists. Each SF Soldier completed initial language training prior to receiving his SF unit assignment, was subsequently assigned to an SF unit and a team (i.e., important for multilevel research), completed a proficiency assessment (i.e., maintenance) as part of required annual testing, and received a rating from his team leader assessing language-related performance in the field (i.e., generalization). This provided a unique opportunity to study two operationalizations of language skill transfer in a team-based work context, since there is rarely systematic measurement of generalization (i.e., performance) in the U.S. military.

Initial Skill Acquisition and Language Transfer

Trainees who demonstrate a higher posttraining competency in acquired skills should, over time, be better equipped to maintain their skills, and to generalize them to different settings. Skill is believed to be one of several key determinants of job performance (Campbell, McCloy, Oppler, & Sager, 1993), and meta-

analytic evidence from the training literature indicates a strong relationship between skill acquisition and transfer (Colquitt, LePine, & Noe, 2000). However, with a few recent exceptions (Dierdorff & Surface, 2008; Ross et al., 2011), little research has investigated the relationship between posttraining foreign language skill and transfer with language-enabled military personnel.¹ With regard to skill maintenance, Dierdorff and Surface (2008) found that initial language proficiency was correlated at .90 or higher with four subsequent measures of language proficiency. However, to our knowledge, the only empirical evidence available directly examining the relationship between language proficiency and job performance (generalization) comes from studies of expatriate performance within civilian populations, studies of foreign language teachers in educational settings, and studies for a small number of other civilian occupations such as call center employees (Mol, Born, Madde, & van der Molen, 2005; Surface, Gissel, & Borneman, 2013). Although many of these studies used self-report measures of language proficiency as well as overall job performance measures, it should be noted that a subsample ($k = 10$) of studies indicated a moderately strong relationship ($r = .42$) between proficiency measures and supervisor ratings of language-related performance (Surface et al., 2013). Given these previous findings, we expect the extent to which trainees have acquired skills from training to be positively related to both skill maintenance and generalization.

Hypothesis 1a: Initial language skill has a positive relationship with skill maintenance.

Hypothesis 1b: Initial language skill has a positive relationship with skill generalization.

Posttraining Time Interval and Language Transfer

Despite the expected positive relationship between initial skill acquisition and transfer, previous research indicates that trained skills are likely to decay over time (Arthur, Bennett, Stanush, & McNelly, 1998; Wang, Day, Kowollik, Schuelke, & Hughes, 2013). However, other researchers have found evidence that under certain circumstances transfer effects can actually

increase over time (Taylor et al., 2005). The authors interpret this finding to be a result of the types of skills examined in their study (i.e., primarily interpersonal communication skills). Moreover, other research also indicates that the degree of decay is influenced by characteristics of the task-content demands, with moderate cognitive/low physical tasks showing the most decay (Wang et al., 2013). Therefore, the nature of the skill/task seems to play an important role in skill decay/retention over time. With regard to foreign language skills in particular, given that communication in a foreign language is a distinctly cognitive skill, we expect a negative relationship between the posttraining time interval and skill maintenance. Furthermore, with the decline of language skills over time, the performance of language-related job tasks should suffer as well; thus, we expect the time interval to have a negative association with skill generalization.

Hypothesis 2a: The posttraining time interval has a negative relationship with skill maintenance.

Hypothesis 2b: The posttraining time interval has a negative relationship with skill generalization.

Team Context and Language Transfer

Job performance and transfer activities happen in context, which represents “situational opportunities and constraints that affect the occurrence and meaning of organizational behavior as well as functional relationships between variables” (Johns, 2006, p. 386). More specifically, dimensions of the discrete context include the task, social, and physical context (Johns, 2006). With regard to the social context in particular, existing research examining situational influences on transfer suggests that the social context and environment can affect the likelihood and extent of trainee transfer (e.g., Rouiller & Goldstein, 1993; Tracey, Tannenbaum, & Kavanagh, 1995). Although much of the existing research on context has focused on groups at higher levels (e.g., units or organizations), teams also represent salient situational

¹ We are using the Department of Defense distinction between linguists (core focus of job) and language-enabled (enhances job) personnel.

influences that may shape individual perceptions and behaviors regarding the maintenance and generalization of trained skills. And, as noted by Porter, 2008 (p. 159), “given the saliency of work groups and teams, it is likely that such effects are stronger than those from higher levels.” Indeed, although research has not examined the impact of the team context on individual language skill transfer, there is some existing evidence suggesting that the team context can facilitate or hinder the transfer of training (Smith-Jentsch et al., 2001). Given the saliency of the demonstration of foreign language skills within the immediate work team for SF members, we expect significant variability in language skill maintenance and generalization associated with teams.

Hypothesis 3a: Teams account for significant variability in trainees’ skill maintenance.

Hypothesis 3b: Teams account for significant variability in trainees’ skill generalization.

Team Mean Posttraining Language Skill

If skill maintenance and generalization vary across team contexts as expected, the question remains as to what team characteristics explain or drive this between-team variation. One important dimension of the social context is the influence that peers have on trainees. In most cases, prior research has examined the extent to which peer support has a positive influence on training transfer (e.g., Tracey et al., 1995). However, another potentially important aspect of peer influence includes the skill-level of peers. According to the theory of planned behavior, subjective norms are one of three key factors (along with attitudes and perceived behavioral control) which influence a person’s intentions and subsequent behavior (Ajzen, 1991). Subjective norms refer to the perceived social pressure to perform or not perform a behavior (Ajzen, 1991), and meta-analytic evidence confirms the importance of subjective norms in shaping behavioral intentions (Armitage & Conner, 2001). Descriptive norms represent a particular type of norm that influences intentions, through which “the opinions and actions of significant others provide information that people may use in deciding what to do themselves (e.g., ‘If everyone’s doing it, then it must be a sensible thing to do,’ cf., Cialdini,

Kallgren, & Reno, 1991)” (Rivis & Sheeran, 2003, p. 220).

This research suggests that, when working in teams, the skill-level of teammates could influence others to be more/less likely to maintain and transfer the newly trained skill. This may be especially true if the skill is one that is practiced or applied saliently through interacting with teammates. For example, a trainee that is surrounded by peers with higher levels of foreign language skills may be more likely to practice that skill to “keep up” or to avoid being perceived poorly. On the contrary, if a trainee is surrounded by peers of comparatively lower levels of skill, then he may be less inclined to practice or utilize these skills (Ajzen, 1991; Armitage & Conner, 2001; Rivis & Sheeran, 2003). Thus, teams with higher/lower initial language skill should, on average, tend to have higher/lower subsequent skill maintenance and generalization.

Hypothesis 4a: Team mean initial language skill has a positive relationship with skill maintenance.

Hypothesis 4b: Team mean initial language skill has a positive relationship with skill generalization.

In addition to the potential direct effects exerted by team characteristics, multilevel models of training effectiveness often posit cross-level moderating influences as well (e.g., Kozlowski et al., 2000; Tracey et al., 1995). In this case, team-level characteristics, such as the mean level of language skill, may act to regulate the associations among trainee characteristics (e.g., trainee initial skill, time interval, etc.) and skill maintenance and generalization. In other words, the relationships between the individual-level predictors and training transfer may differ for teams with low versus high average levels of language skill. Despite sound theoretical support for such effects, Kozlowski et al., 2000 (p. 167) noted that, “Support for this moderating effect has been more elusive.” For example, Tracey et al. (1995) found a direct effect for the transfer climate, however no such evidence was revealed for a moderating effect. In contrast, recent research by Ellington and Dierdorff (2014) found that team context factors moderated the relationship between individual self-regulation during training and subsequent learning transfer to a novel task. Given the mixed support for cross-

level moderation from previous research, in this study we take an exploratory approach to investigating these relationships.

Research Question 1: Do the relationships between the individual predictors and language skill maintenance and generalization vary across teams?

Research Question 2: Does the team mean initial language skill moderate the relationships between the individual predictors and language skill maintenance and generalization?

Method

Participants

Participants in this study consisted of U.S. Army Special Forces (SF; Green Berets) personnel who previously completed job-required initial acquisition foreign language training during the Special Forces Qualification Course (SFQC), prior to being assigned to an operational unit.² SF personnel are required to learn and maintain a language in order to enhance their core mission activity. They are considered language-enabled personnel, not linguists, since language is not their core mission. More specifically, 919 active-duty SF Soldiers from 133 Operational Detachments Alpha (ODA; i.e., teams) had a posttraining language proficiency test score, had a subsequent valid language proficiency score, and received a language performance rating from their team leader. All of the participants were male noncommissioned officers. The sample covered the enlisted military occupational specialties on a SF team. Many of these occupations are similar to occupations found in other military organizations or in civilian organizations, such as the SF medic (18D), who must earn and maintain the same certifications as an emergency medical technician, in addition to SF training (e.g., small unit tactics). The foreign language performance ratings in our study are unique in the military since language performance is not measured typically, and these ratings were obtained as part of an organizational research study of SF field performance (Wilson, Drewes, Cunningham, Sanders, Thompson, & Surface, 2001; Wilson & Sanders, 2003). Therefore, SF Soldiers in our sample were participants in the SF field performance study, conducted by North Carolina State University and the U.S. Army Research

Institute. The study was conducted for the US Army Special Forces Command and U.S. Army John F. Kennedy Special Warfare Center and School (USAJFKSWCS). These data came from this research archive and were integrated with training and testing data from a project sponsored by the Special Operations Forces Language Office, U.S. Special Operations Command (Donnelly & Surface, 2006).

Procedure

Although this study uses archival data, the sample context is known. Foreign language training took place over 18–24 weeks at USAJFKSWCS, depending on the difficulty of the language for a native English speaker. The course curriculum was standardized across all training courses/languages, including materials and learning objectives. The training was not SF or military specific since it was focused on general language proficiency. At this time, the language training was very similar to other military and civilian language schools. Trainees' skill in the target language was assessed upon completion of initial acquisition training using an official standardized general proficiency test in nonparticipatory listening and reading with all languages having the same graduation standard. Language skill maintenance (using the same official proficiency test) and generalization were then subsequently assessed at varying posttraining time intervals, which depended on a variety of factors, such as deployments and test availability.

Measures

Initial language skill and maintenance. Initial posttraining language skill, as well as subsequent skill maintenance, was assessed using the Defense Language Proficiency Test (DLPT), which was the test of record at the time of this research.³ Designed by the Defense Language Institute (DLI), the DLPT

² The placement of language within the SFQC has varied but was always prior to ODA assignment.

³ The DLPT version in this study is DLPT IV. Most languages now are tested with DLPT 5. Although USSOCOM still uses the DLPT, the official test of record at writing is the Oral Proficiency Interview (OPI). The OPI is an interview-formatted assessment designed to elicit a ratable sample of speech that can be rated on the ILR scale. Speaking is the primary language skill used by SF (SWA Consulting Inc., 2010; 2013).

measures both nonparticipatory listening and reading skill proficiency, and is based on the Interagency Language Roundtable (ILR) framework, which defines general proficiency in terms of functional skill levels, ranging from ILR Level 0 (no proficiency) to ILR Level 5 (*Functionally Native* proficiency). The DLPT is specific to the language trained, and includes a 2.5 hour reading comprehension assessment, and a 1.5 hour listening comprehension assessment (Silva & White, 1993). The DLPT ranges from ILR Level 0 to ILR Level 3 (*General Professional* proficiency). For analysis purposes, the ILR values were recoded into a 7-point scale, with higher values indicating higher levels of language skill proficiency. In addition, for research purposes, a DLPT composite score was computed for each participant by averaging the listening and reading component scores (Dierdorff & Surface, 2008). Finally, team (i.e., ODA) mean language skill was operationalized as the average initial posttraining DLPT composite score among all members of a team.

Language skill generalization. The extent to which foreign language training generalized to the performance of job related tasks was measured via performance ratings collected from each participant's Team Leader (i.e., officer). The appraisal instrument used was specifically developed for the SF (Thompson, 2001), and was based on a previous job analysis which identified 15 key performance categories common to all SF Soldiers, one of which was "Using and Enhancing Language Skills" (Russell, Crafts, Tagliarini, McCloy, & Barkley, 1996). Behavioral statements taken directly from Russell et al.'s (1996) job analysis were incorporated into a mixed standard rating scale (MSRS) format (Blanz & Ghiselli, 1972), with several of these behaviors (i.e., picks up languages readily; uses language skillfully; translates adeptly, rarely, if ever, miscommunicating information) comprising a language performance dimension. The MSRS required raters to indicate the ratee's performance on each behavior using one of three alternatives: better than (+), less than (-), or equal to (0) the behavioral examples representing varying levels of performance effectiveness (i.e., low, average, and high). The internal consistency of the three items comprising the language performance dimension was acceptable ($\alpha = .71$). Ratings on these behaviors

were combined using Saal's algorithm (1979) to derive a dimensional rating for language performance on a 7-point scale, with higher values indicating higher performance. These data were collected during an organizational research initiative.

Time interval. The posttraining language skill maintenance time interval was operationalized as the number of months between the completion of initial acquisition training and the second language skill assessment. Skill maintenance assessment time intervals ranged from a minimum of three months posttraining to a maximum of 88 months ($M = 13.69$, $SD = 8.00$). The posttraining language skill generalization time interval was measured as the number of months between initial acquisition training and the subsequent performance evaluation. Skill generalization assessment time intervals ranged from a minimum of three months posttraining to a maximum of 49 months ($M = 20.99$, $SD = 7.50$).

Control variables. General cognitive ability is known to predict successful learning and job performance (Ree, Carretta, & Teachout, 1995; Ree & Earles, 1991; Schmidt & Hunter, 1998); therefore, it was useful to examine the relationships among our substantive variables of interest while holding this individual difference variable constant. Cognitive ability was measured using the Armed Forces Qualification Test (AFQT), which is a composite score based on several subsections of the Armed Services Vocational Aptitude Battery (ASVAB). Although item-level data was not available for the current sample, previous research has suggested high reliability for the AFQT (i.e., .94; "Official Site of the ASVAB Testing Program," 2014).

In addition, because participants were trained in a variety of languages, it was important to hold the difficulty of the task constant when examining our primary hypotheses. Similar to prior research (Dierdorff & Surface, 2008), task (i.e., language) difficulty was operationalized using the four-category classification system used by the U.S. government. The categories reflect the increasing difficulty for a native English speaker to learn the language (Silva & White, 1993). For example, Spanish is a Category I language, German is a Category II, Russian is a Category III, and Arabic is a Category IV. Task difficulty was coded 1–4, with higher levels indicating greater difficulty.

Analytical Strategy

In order to address our hypotheses and research questions, a multilevel modeling approach was employed, with trainees (i.e., SF Soldiers) comprising level-1 and teams (i.e., ODAs) comprising level-2. A staged modeling approach (Aguinis, Gottfredson, & Culpepper, 2013; Hox, 1995; Raudenbush & Bryk, 2002) was followed for each of the two transfer criterion variables. The first stage included the estimation of unconditional or “null” models with no predictors, in order to partition the within and between-team variance in the outcomes. Second, random intercept and fixed slope models (RIFSM) were estimated, which allowed intercepts to vary across teams, and included the level-1 predictors with fixed slopes. These models also included the level-2 main effect for team mean language skill. Third, random intercept and random slope models (RIRSM) were estimated in order to examine whether significant variance existed in the relationships (i.e., slopes) between the level-1 predictors and criteria. Following the recommendations of Hox (1995), random slope variation was initially tested one predictor at a time, with a subsequent model including all variance components that were identified as significant when tested individually. Finally, providing significant variation in any of the level-1 slopes, cross-level interaction models (CLIM) were estimated in order to examine potential moderating effects of team mean language skill on the lower-level relationships. In the skill maintenance models, the level-1 predictors were centered around their grand mean. It was therefore necessary to estimate the

level-2 interaction between team mean language skill and team mean task difficulty in addition to the cross-level effect, since this is recommended to avoid spurious cross-level interactions (Aguinis et al., 2013). In modeling skill generalization, given the potential for social comparison effects in performance evaluations within teams, the level-1 predictors were centered around their group means, which precludes the need to control for across-group variance (Aguinis et al., 2013).

Results

Table 1 presents descriptive statistics and zero-order correlations for all study variables. The multilevel modeling results for language skill maintenance are presented in Table 2, with the RIFSM results indicating that initial language skill ($\gamma_{30} = .85, p < .01$) was a significant positive predictor of skill maintenance, supporting Hypothesis 1a. In support of Hypothesis 2a, the posttraining time interval ($\gamma_{40} = -.01, p < .01$) was also significant, with longer time intervals associated with lower skill maintenance scores. The null model results provide support for Hypothesis 3a, given the significant between-team variability in skill maintenance ($\tau_{00} = 1.28, df = 132, \chi^2 = 744.82, p < .001$), with approximately 40% of the variance residing between teams. In addition, the team mean initial language skill predicted significant between-team variance in skill maintenance ($\gamma_{01} = .16, p < .01$), providing support for Hypothesis 4a. A model comparison test indicated a significant improvement in fit over the null model ($\chi^2 = 1418.15, p < .001$), and the collective set of

Table 1
Descriptive Statistics and Zero-Order Correlations

	<i>M</i>	<i>SD</i>	1	2	3	4	5	6
Trainee-level (L1) variables								
Cognitive ability	221.23	20.81						
Task difficulty	2.19	1.16	.14**					
Initial language skill	3.41	1.65	.03	-.44**				
Time interval (skill maintenance)	13.69	8.00	-.05	.19**	-.15**			
Time interval (skill generalization)	20.99	7.50	.03	-.05	.09**	.01		
Language skill maintenance	3.29	1.81	.00	-.54**	.90**	-.21**	.15**	
Language skill generalization	5.74	1.37	.07*	-.01	.29**	-.01	.04	.28**
Team-level (L2) variables								
Team mean initial language skill	3.21	1.60						

Note. L1 = Level 1; L2 = Level 2; L1 *N* = 919 and L2 *N* = 133.

* $p < .05$. ** $p < .01$.

Table 2
Multilevel Modeling Results for Transfer as Language Skill Maintenance

Level and variable	Model			
	Null	Random intercept and fixed slope	Random intercept and random slope	Cross-level interaction
Trainee-level (L1)				
Intercept (γ_{00})	3.260** (.109)	3.282** (.028)	3.284** (.027)	3.224** (.034)
Cognitive ability (γ_{10})		-.000 (.001)	.000 (.001)	.000 (.001)
Task difficulty (γ_{20})		-.193** (.034)	-.186** (.034)	-.219** (.037)
Initial language skill (γ_{30})		.851** (.018)	.849** (.018)	.848** (.018)
Time interval (γ_{40})		-.011** (.003)	-.011** (.003)	-.012** (.003)
Team-level (L2)				
Team mean ILS (γ_{01})		.155** (.038)	.163** (.038)	.105** (.041)
Team mean ILS \times team mean task difficulty (γ_{02})				-.014 (.081)
Cross-level interaction				
Team mean ILS \times task difficulty (γ_{21})				-.077 (.068)
Variance components				
Within-team (L1) variance (σ^2)	1.963	.512	.507	.506
Intercept (L2) variance (τ_{00})	1.279**	.026**	.006*	.003*
Task difficulty slope (L2) variance (τ_{11})			.019*	.017*
Additional information				
ICC	.395			
-2 Log likelihood (FIML)	3,449	2,030**	2,026	2,018*
Number of estimated parameters	3	8	10	12
Within-team (L1) pseudo R^2		.739	.742	.742
Intercept (L2) pseudo R^2		.979	.996	.997
Task difficulty slope (L2) pseudo R^2				.116

Note. L1 = Level 1; L2 = Level 2; L1 $N = 919$ and L2 $N = 133$; ILS = initial language skill; ICC = intraclass correlation coefficient; FIML = full information maximum likelihood estimation; L1 predictors are centered around their grand mean; values in parentheses are robust standard errors; t -statistics were computed as the ratio of each regression coefficient divided by its standard error.

* $p < .05$. ** $p < .01$.

predictors explained 74% of the within-team variance and 98% of the between-team variance in skill maintenance. In addressing Research Question 1, the RIRSM then tested for slope variation in the level-1 predictors, and found that the task difficulty slope varied significantly across teams ($\tau_{11} = .02$, $df = 79$, $\chi^2 = 105.22$, $p < .05$). However, the model comparison did not indicate a significant improvement in fit over the RIFSM ($\chi^2 = 4.20$, $p = .12$). Because there was some evidence of slope variation for task difficulty, a CLIM was examined to address Research Question 2 and determine if team mean language skill moderates the task difficulty relationship with skill maintenance, however the results did not suggest cross-level moderation ($\gamma_{21} = -.08$, $p > .05$). The final model comparison did, nevertheless, show an improvement in fit over the RIRSM ($\chi^2 = 8.03$, $p < .05$), suggesting somewhat inconclusive

results regarding potential cross-level moderation for skill maintenance. Finally, the CLIM results indicated that the within-team variance explained remained at 74%, with 99% of the between-team intercept variance explained.

Table 3 presents the multilevel modeling results for skill generalization to job performance. The RIFSM results provided support for Hypothesis 1b in that initial language skill was a significant predictor ($\gamma_{30} = .33$, $p < .01$), however the post-training time interval ($\gamma_{40} = .33$, $p > .05$) was not associated with skill generalization, therefore failing to support Hypothesis 2b. Null model results suggested significant between-team variance in skill generalization ($\tau_{00} = .41$, $df = 132$, $\chi^2 = 387.58$, $p < .001$), with roughly 22% of the variability associated with teams, thus supporting Hypothesis 3b. Team mean language skill did not explain significant between-team variability in skill generalization ($\gamma_{01} = .09$, $p > .05$), provid-

Table 3
Multilevel Modeling Results for Transfer as Language Skill Generalization to Job Performance

Level and variable	Model			
	Null	Random intercept and fixed slope	Random intercept and random slope	Cross-level interaction
Trainee-level (L1)				
Intercept (γ_{00})	5.747** (.069)	5.747** (.069)	5.748** (.069)	5.749** (.069)
Cognitive ability (γ_{10})		.005* (.002)	.004* (.002)	.005* (.002)
Task difficulty (γ_{20})		.125* (.064)	.091 (.061)	.087 (.061)
Initial language skill (γ_{30})		.328** (.040)	.320** (.038)	.295** (.036)
Time interval (γ_{40})		-.001 (.005)	.001 (.005)	.001 (.005)
Team-level (L2)				
Team mean ILS (γ_{01})		.094 (.066)	.193** (.054)	.099 (.066)
Cross-Level Interaction				
Team mean ILS \times ILS (γ_{31})				.120** (.039)
Variance components				
Within-team (L1) variance (σ^2)	1.471	1.118	1.094	1.098
Intercept (L2) variance (τ_{00})	.410**	.432**	.466**	.456**
ILS Slope (L2) variance (τ_{11})			.083**	.066**
Additional information				
ICC	.218			
-2 Log likelihood (FIML)	3,101	2,971**	2,908**	2,898**
Number of estimated parameters	3	8	10	11
Within-team (L1) pseudo R^2		.150	.256	.254
Intercept (L2) pseudo R^2		.000	.000	.000
ILS slope (L2) pseudo R^2				.209

Note. L1 = Level 1; L2 = Level 2; L1 $N = 919$ and L2 $N = 133$; ILS = initial language skill; ICC = intraclass correlation coefficient; FIML = full information maximum likelihood estimation; L1 predictors are centered around their group mean; values in parentheses are robust standard errors; t -statistics were computed as the ratio of each regression coefficient divided by its standard error.

* $p < .05$. ** $p < .01$.

ing no support for Hypothesis 4b. The RIFSM showed improved model fit as compared to the null model ($\chi^2 = 129.83$, $p < .001$), with the predictors explaining 15% of the within-team variance, but none of the between-team variance. In addressing Research Question 1, the RIRSM results suggested that the relationship between initial language skill and subsequent skill generalization varied across teams ($\tau_{11} = .08$, $df = 132$, $\chi^2 = 237.50$, $p < .001$). In addition, a model comparison showed an improvement in fit as compared to the RIFSM ($\chi^2 = 63.07$, $p < .001$). With regard to Research Question 2, given evidence of significant slope variability for language skill, a CLIM was then examined, which indicated that team mean language skill predicted variability in the individual language skill slopes ($\gamma_{31} = .12$, $p < .01$). Figure 1 depicts this cross-level moderation effect, and shows a stronger positive association between language skill and skill generalization in teams with higher mean language skill. The CLIM also showed improved model fit as com-

pared to the RIRSM ($\chi^2 = 10.68$, $p < .001$). The predictors explained 25% of the within-team variance, and although the variables did not explain significant intercept variability, team mean language skill explained 21% of the between-team language skill slope variability.

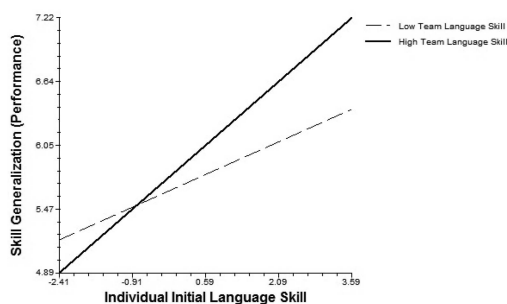


Figure 1. Cross-level moderating effect of team mean initial language skill on the relationship between individual initial language skill and skill generalization (performance).

Discussion

Military organizations invest tremendous resources in training their personnel in order to ensure effective performance when called upon. Therefore, identifying and understanding the factors related to the transfer of trained skill to work contexts is critical for maximizing individual, team, and organizational performance. Our results indicate that, although there are certainly some similarities in the findings for skill maintenance and generalization, there are also differences worth noting, which suggest important considerations and implications for the measurement of transfer, as well as military policy surrounding language-related performance, especially for Special Operations and other language-enabled forces (i.e., nonlinguists). Below we review our findings and discuss practical implications and suggestions for future research throughout.

Individual-Level Findings

Transfer of initial posttraining skill. In reviewing our individual-level findings, posttraining language skill was found to be a significant predictor of both skill maintenance and skill generalization transfer measures. This indicates that those who learn more in training are better able to apply/utilize their language skills within the job context, which is an encouraging confirmation that had not been demonstrated in prior language training research. Initial language skill was a stronger predictor of skill maintenance ($r = .90$) than it was of skill generalization ($r = .29$), however this is not surprising as the U.S. military's skill maintenance approach uses the same foreign language proficiency assessment repeated annually, and the same standardized assessment repeated should be highly correlated over measurement occasions. The skill maintenance finding is consistent with the correlation of .92 between the initial language skill assessment and the first maintenance assessment reported by [Dierdorff and Surface \(2008\)](#). With regard to skill generalization, the correlation between posttraining skill and transfer found here was also not as strong as that reported in previous meta-analytic research (i.e., .50; [Colquitt et al., 2000](#)). This indicates that initial skill assessments had a

more moderate relationship with the application of this skill on the job.

There are several possible explanations as to why our results suggest a moderate relationship between tested language skill and language-related field performance. First, from a theoretical perspective, performance models suggest that skill is but one of several determinants of job performance, with others including declarative knowledge and motivation ([Campbell et al., 1993](#)). For example, it is possible that despite adequate skill and motivation to perform on proficiency tests (e.g., due to skill-based compensation), insufficient motivation to apply these skills on the job may attenuate the association between posttraining capabilities and transfer as performance. Second, our results may be due to a misalignment between the skill proficiency test and most salient communication aspects of the SF mission. At the time of this research, the assessment of language skill for SF focused on nonparticipatory listening and reading (i.e., DLPT), which are skills required for signals intelligence. However, the most salient aspects of SF missions, such as conversing (e.g., building rapport), giving commands (e.g., directing) and presenting (e.g., training foreign personnel), require speaking and participatory listening skills ([Surface, Poncheri, Lemmond, & Shetye, 2005](#)). Future research should explore both of the above possibilities in order to determine the extent to which Soldiers are motivated to apply their language skills and to ensure alignment between tested skills and job requirements. For example, a thorough needs assessment is recommended to align training, assessment, and policy with capability requirements for individual and organizational performance ([Surface, 2012](#)).

Since the project that served as the basis for this study was completed, U.S. Special Operations Command (USSOCOM) has aligned training, testing, and policy more with the mission capability requirements of speaking, as several studies have documented the speaking requirements of SF ([SWA Consulting Inc., 2010; 2013](#)). For example, the language proficiency test of record for SF was changed from the DLPT (i.e., measures nonparticipatory listening and reading) to a measure of speaking proficiency, the Oral Proficiency Interview (OPI; [United States Special Operations Command, 2009](#)). In this case, the OPI is an assessment

conducted by a trained interviewer over the phone who elicits a sample of speech that is recorded and then rated by the interviewer and another rater independently. Recently, the OPI was added to the promotion requirements for SF Non-Commissioned Officers (NCOs; [Department of the Army, 2012](#)). It should be noted that both the DLPT and OPI are general proficiency measures, not contextualized to specific SF language requirements. A more contextualized skill measure might transfer more effectively to SF mission contexts. There have been recent calls in the Special Operations Forces (SOF) language community to focus on capability-based language assessment that is derived from a thorough analysis of the mission requirements instead of a general model of language proficiency ([Federe, 2014](#); [Surface, 2014](#)). Future research should investigate the skill-performance relationship with the OPI as the skill maintenance measure, and also revisit studies such as [Dierdorff and Surface \(2008\)](#), which also used the reading and nonparticipatory listening assessment as their criterion with SF.

Finally, a key practical question evoked by our results is whether skill maintenance—focused on developing and maintaining sufficient skill necessary for performance—provides sufficient information alone regarding the transfer of foreign language training. Our findings suggest that future research addressing this issue would be relevant to other communities of interest—both military and nonmilitary. The General Purpose Forces are deployed outside the continental U.S. and must be able to communicate and interact with local populations from numerous regions all over the world while conducting a variety of missions and tasks, ranging from searching a building to conducting a medical clinic for a village. There are a number of military occupations for which language is considered an enhancing or enabling skill. In addition, expatriates that accept multinational assignments must also learn and maintain appropriate language skills and perform them to accomplish a variety of tasks on the job and in daily life. Language is critically important for building rapport, whether it is an emergency responder in South Texas speaking with a Spanish-speaking grandmother, an SF Soldier speaking with the village elders in Iraq, or a businessperson negotiating a deal in Seoul, Korea. Is a measure of skill maintenance, especially a general proficiency measure, sufficient to under-

stand if an individual can perform in context? Though there are costs associated with additional on-the-job assessments such as supervisory or peer evaluations of language-related performance (e.g., time of key personnel, access to personnel while on assignment), our findings suggest that skill maintenance and skill generalization can both have value in understanding the effectiveness of training and the readiness of individuals to perform their jobs and missions.

Transfer time interval. As trained skills often decay over time ([Arthur et al., 1998](#); [Wang et al., 2013](#)), we expected time intervals between the posttraining skill (initial acquisition) and the transfer measurements to be negatively associated with both skill maintenance and skill generalization. However, the hypothesized relation was found only for skill maintenance and not for skill generalization. There are several possible explanations for this finding. For example, it may be that, despite an actual decay in proficiency as measured using the non-contextualized skill assessment, the effectiveness with which Soldiers perform the contextualized language-related job tasks remains more consistent. In addition, the differential results may be a function of the nature of the measures and measurement contexts. Transfer as maintenance was operationalized using the same standardized high-stakes individual assessment (i.e., DLPT) for both initial skill and skill maintenance. This type of measurement context creates a “maximum” transfer context, in which trainees are aware that they are being evaluated and are likely to maximize effort ([Huang, Blume, Ford, & Baldwin, 2012](#)). Transfer as generalization was operationalized using supervisory ratings, which can be viewed as a more “typical” transfer measure, and stems more from a trainee’s volition to apply their skills to work tasks over time ([Huang et al., 2012](#)).

In order to better understand the impact of posttraining time intervals, future research should capture relevant events occurring during the interval (e.g., additional language training), as well as contextual factors (e.g., incentives) that might impact transfer. In addition, since our findings suggest that language skill tended to decay over time, more research is needed to develop an optimal recertification interval and guide the timing of refresher training for language-enabled personnel.

Team-Level Findings

Our study also examined the extent to which the context created by work teams (i.e., ODAs) influence the transfer of foreign language skills. Our null model results supported our hypothesis that teams would account for significant variability in both skill maintenance and generalization. However, the findings from our subsequent model results offer different interpretations of the team effects for our two transfer criteria. More specifically, though initial model results indicated that approximately 40% of the variability in skill maintenance was associated with teams, including individual differences and other predictors explained almost all (98%) of the between-team variance. This suggests that, although there is variance across teams in skill maintenance, it seems to be primarily due to systematic differences across the teams in terms of initial language skill, posttraining time intervals, cognitive ability, and task (language) difficulty. However, teams that had higher/lower initial language skill tended to have higher/lower skill maintenance scores. Interestingly, the relationship between language difficulty and skill maintenance varied across teams; however, this variability was not associated with the team mean language skill. Future research should investigate other team context factors that might explain this variability.

The null model results also supported our prediction regarding teams and skill generalization, with 22% of the variance associated with work teams. In addition, the relationship between individual initial language skill and generalization varied across teams, and the team mean skill level moderated this association such that there was a stronger relationship with performance in teams of higher average skill. Previous research suggests that aspects of the work environment such as a supportive work climate can enhance training transfer (Tracey et al., 1995), and our findings indicate that the skill-level of one's peers can also be influential. Ours is the first study that we are aware of that has tested such effects. Therefore, future research is needed to replicate our results and further investigate explanatory mechanisms and boundary conditions. For example, we speculated that one reason that peer skill may be influential is through the establishment of norms which shape behavior (Ajzen, 1991; Armitage & Con-

ner, 2001), thus research is needed to determine if these norms are indeed driving the direct or moderating influence of team average skill. From a practical standpoint, if evidence could be found to support these propositions, military organizations and others could potentially use individual skill levels to inform team formation and to shape norms for skill maintenance and application.

Limitations

Our findings should be interpreted in light of the limitations of this research. First, given the use of archival data, we were unable to exercise experimental control, or explore other individual and team characteristics that may be important in shaping the maintenance and generalization of language skill. In addition, based on our data, it is difficult to determine the extent to which our finding of team-level variability in skill generalization is due to characteristics of the team context, or a measurement artifact due to "rater effects" in performance ratings (Hoffman, Lance, Bynum, & Gentry, 2010; O'Neill, Goffin, & Gellatly, 2012). Finally, the extent to which our findings generalize to other populations, occupations, or skills is unknown; therefore, further research within military and other organizations is needed. However, there are reasons to believe that our findings will generalize to other language training contexts since the SF foreign language training at the time of this study was fairly consistent with other military organizations and with university training, and there is some occupational task overlap with other military and some civilian jobs.

Conclusion

The U.S. military currently uses a "skill maintenance" strategy for the assessment of foreign language transfer and for related administrative purposes (e.g., foreign language proficiency bonus). Although skill maintenance is easier from an administrative perspective, it must be acknowledged that the U.S. military is paying for maintaining a skill at a certain level, not its application on the job. If the skill application on the job is important, then a skill generalization measure must be used. This research can be used to inform such policy decisions in the future. From a research perspective, this

study can also inform both the language and transfer research literatures. We encourage more research investigating multiple definitions of transfer as well as individual and contextual influences.

References

- Aguinis, H., Gottfredson, R. K., & Culpepper, S. A. (2013). Best-practice recommendations for estimating cross-level interaction effects using multilevel modeling. *Journal of Management*, 39, 1490–1528. <http://dx.doi.org/10.1177/0149206313478188>
- Aguinis, H., & Kraiger, K. (2009). Benefits of training and development for individuals and teams, organizations, and society. *Annual Review of Psychology*, 60, 451–474. <http://dx.doi.org/10.1146/annurev.psych.60.110707.163505>
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50, 179–211. [http://dx.doi.org/10.1016/0749-5978\(91\)90020-T](http://dx.doi.org/10.1016/0749-5978(91)90020-T)
- Armitage, C. J., & Conner, M. (2001). Efficacy of the theory of planned behaviour: A meta-analytic review. *British Journal of Social Psychology*, 40, 471–499. <http://dx.doi.org/10.1348/014466601164939>
- Arthur, W., Jr., Bennett, W., Jr., Stanush, P. L., & McNelly, T. L. (1998). Factors that influence skill decay and retention: A quantitative review and analysis. *Human Performance*, 11, 57–101. http://dx.doi.org/10.1207/s15327043hup1101_3
- Baldwin, T. T., Ford, J. K., & Blume, B. D. (2009). Transfer of training 1988–2008: An updated review and agenda for future research. In G. P. Hodgkinson & J. K. Ford (Eds.), *International review of industrial and organizational psychology* (Vol. 24, pp. 41–70). Chichester, UK: Wiley, Ltd.
- Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, 128, 612–637. <http://dx.doi.org/10.1037/0033-2909.128.4.612>
- Blanz, F., & Ghiselli, E. E. (1972). The mixed standard scale: A new rating system. *Personnel Psychology*, 25, 185–199. <http://dx.doi.org/10.1111/j.1744-6570.1972.tb01098.x>
- Blume, B. D., Ford, J. K., Baldwin, T. T., & Huang, J. L. (2010). Transfer of training: A meta-analytic review. *Journal of Management*, 36, 1065–1105. <http://dx.doi.org/10.1177/0149206309352880>
- Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1993). A theory of performance. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 35–70). San Francisco, CA: Jossey-Bass.
- Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. In P. Z. Mark (Ed.), *Advances in experimental social psychology: Vol. 24* (pp. 201–234). New York, NY: Academic Press. [http://dx.doi.org/10.1016/S0065-2601\(08\)60330-5](http://dx.doi.org/10.1016/S0065-2601(08)60330-5)
- Colquitt, J. A., LePine, J. A., & Noe, R. A. (2000). Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology*, 85, 678–707. <http://dx.doi.org/10.1037/0021-9010.85.5.678>
- Department of the Army. (2012). *U.S. Army commissioned officer professional development guide (Pamphlet 600–3)*. Washington, D. C.
- Dierdorff, E. C., & Surface, E. A. (2008). If you pay for skills, will they learn? Skill change and maintenance under a skill-based pay system. *Journal of Management*, 34, 721–743. <http://dx.doi.org/10.1177/0149206307312507>
- Donnelly, J., & Surface, E. A. (2006, February). *Language Proficiency Testing and Special Forces Field Performance*. A presentation to the Defense Language Action Panel. Washington, DC.
- Ellington, J. K., & Dierdorff, E. C. (2014). Individual learning in team training: Self-regulation and team context effects. *Small Group Research*, 45, 37–67. <http://dx.doi.org/10.1177/1046496413511670>
- Federe, M. (2014). *Developing a mission-focused, capability-based language curriculum for naval special warfare*. Paper presented at the Interagency Language Roundtable (ILR) Plenary Session, National Foreign Language Center, MD.
- Hackman, J. R. (2003). Learning more by crossing levels: Evidence from airplanes, hospitals, and orchestras. *Journal of Organizational Behavior*, 24, 905–922. <http://dx.doi.org/10.1002/job.226>
- Hoffman, B., Lance, C. E., Bynum, B., & Gentry, W. A. (2010). Rater source effects are alive and well after all. *Personnel Psychology*, 63, 119–151. <http://dx.doi.org/10.1111/j.1744-6570.2009.01164.x>
- Hox, J. J. (1995). *Applied multilevel analysis*. Amsterdam: TT-Publikaties.
- Huang, J. L., Blume, B. D., Ford, J. K., & Baldwin, T. T. (2012). *Paths to transfer: A meta-analytic investigation of training outcomes*. In M. Wang & L. Zhou (Chairs), New developments in training motivation and training transfer research. Symposium conducted at the 27th Annual Conference for the Society for Industrial and Organizational Psychology, San Diego, CA.
- Johns, G. (2006). The essential impact of context on organizational behavior. *The Academy of Management Review*, 31, 386–408. <http://dx.doi.org/10.5465/AMR.2006.20208687>

- Kozlowski, S. W. J., Brown, K. G., Weissbein, D. A., Cannon-Bowers, J. A., & Salas, E. (2000). A multilevel approach to training effectiveness: Enhancing horizontal and vertical transfer. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 157–210). San Francisco, CA: Jossey-Bass.
- Mol, S. T., Born, M. P. H., Madde, E. W., & van der Molen, H. T. (2005). Predicting expatriate job performance for selection purposes: A quantitative review. *Journal of Cross-Cultural Psychology*, *36*, 590–620. <http://dx.doi.org/10.1177/0022022105278544>
- Official Site of the ASVAB Testing Program. (2014). *Test Score Precision*. Retrieved January 11, 2014, from http://official-asvab.com/reliability_res.htm
- O'Neill, T. A., Goffin, R. D., & Gellatly, I. R. (2012). The use of random coefficient modeling for understanding and predicting job performance ratings: An application with field data. *Organizational Research Methods*, *15*, 436–462. <http://dx.doi.org/10.1177/1094428112438699>
- Porter, C. O. L. H. (2008). A multilevel, multiconceptualization perspective of goal orientation in teams. In V. I. Sessa & M. London (Eds.), *Work group learning: Understanding, improving, and assessing how groups learn in organizations* (pp. 149–173). New York, NY: Erlbaum.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Ree, M. J., Carretta, T. R., & Teachout, M. S. (1995). Role of ability and prior knowledge in complex training performance. *Journal of Applied Psychology*, *80*, 721–730. <http://dx.doi.org/10.1037/0021-9010.80.6.721>
- Ree, M. J., & Earles, J. A. (1991). Predicting training success: Not much more than g. *Personnel Psychology*, *44*, 321–332. <http://dx.doi.org/10.1111/j.1744-6570.1991.tb00961.x>
- Rivis, A., & Sheeran, P. (2003). Descriptive norms as an additional predictor in the theory of planned behaviour: A meta-analysis. *Current Psychology: A Journal for Diverse Perspectives on Diverse Psychological Issues*, *22*, 218–233. <http://dx.doi.org/10.1007/s12144-003-1018-2>
- Ross, S. J., Bloomfield, A. N., Masters, M. C., Nielson, K. B., Kramasz, D. M., Connell, S. P. O., & Gynther, K. (2011). How does foreign language proficiency change over time? Results of preliminary data mining. *University of Maryland Center for Advanced Study of Language*. Retrieved September 11, 2012, from <http://www.casl.umd.edu/node/1967>
- Rouiller, J. Z., & Goldstein, I. L. (1993). The relationship between organizational transfer climate and positive transfer of training. *Human Resource Development Quarterly*, *4*, 377–390. <http://dx.doi.org/10.1002/hrdq.3920040408>
- Russell, T. L., Crafts, J. L., Tagliarini, F. A., McCloy, R. A., & Barkley, P. (1996). *Job analysis of Special Forces jobs*. (ARI Research Note 96–76). Alexandria, VA: U.S. Army Research Institute of Behavioral and Social Sciences.
- Saal, F. E. (1979). Mixed standard rating scale: A consistent system for numerically coding inconsistent response combinations. *Journal of Applied Psychology*, *64*, 422–428. <http://dx.doi.org/10.1037/0021-9010.64.4.422>
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262–274. <http://dx.doi.org/10.1037/0033-2909.124.2.262>
- Silva, J. M., & White, L. A. (1993). Relation of cognitive aptitudes to success in foreign language training. *Military Psychology*, *5*, 79–93. http://dx.doi.org/10.1207/s15327876mp0502_1
- Smith-Jentsch, K. A., Salas, E., & Brannick, M. T. (2001). To transfer or not to transfer? Investigating the combined effects of trainee characteristics, team leader support, and team climate. *Journal of Applied Psychology*, *86*, 279–292. <http://dx.doi.org/10.1037/0021-9010.86.2.279>
- Surface, E. A. (2012). Training needs assessment: Aligning learning and capability with performance requirements and organizational objectives. In M. A. Wilson, W. Bennett, Jr., S. G. Gibson, & G. M. Alliger (Eds.), *The handbook of work analysis: Methods, systems, applications and science of work measurement in organizations* (pp. 437–462). New York, NY: Routledge/Taylor & Francis Group.
- Surface, E. A. (2014). *Using needs assessment to align learning and assessment with the capability needed to perform a specific mission*. Paper presented at the Interagency Language Roundtable (ILR) Testing Committee, National Foreign Language Center, MD.
- Surface, E. A., Gissel, A., & Borneman, M. (2013). *A meta-analytic investigation of the relationship between language proficiency and performance*. Paper presented at the Annual Meeting of the Language Testing Research Consortium, Seoul, Korea.
- Surface, E. A., Poncheri, R. M., Lemmond, G., & Shetye, T. (2005). *SOF language transformation strategy needs assessment project: Final project report*. (Tech. Rep. No. 20040606). Raleigh, NC: Surface, Ward, & Associates.
- SWA Consulting, Inc. (2010, October). *US Army Special Forces Language Proficiency Requirements Needs Assessment* (Technical Report #2010010623). Raleigh, NC: Author.

- SWA Consulting, Inc. (2013, September). *USASFC(A) LNA methodology and lessons learned report* (Technical Report #2010 010623). Raleigh, NC: Author.
- Taylor, P. J., Russ-Eft, D. F., & Chan, D. W. L. (2005). A meta-analytic review of behavior modeling training. *Journal of Applied Psychology, 90*, 692–709. <http://dx.doi.org/10.1037/0021-9010.90.4.692>
- Thompson, J. A. (2001). *The relationship between scaled behavioral ratings, performance dimension ratings and rankings using United States Army Special Forces soldiers*. (Master's thesis), North Carolina State University, Raleigh, NC. Retrieved from ProQuest Dissertations & Theses Full Text. (304714796)
- Tracey, J. B., Tannenbaum, S. I., & Kavanagh, M. J. (1995). Applying trained skills on the job: The importance of the work environment. *Journal of Applied Psychology, 80*, 239–252. <http://dx.doi.org/10.1037/0021-9010.80.2.239>
- United States Special Operations Command. (2009). *USSOCOM Manual 350–8: Training: The Special Operations Forces language program*. MacDill Air Force Base, FL.
- U.S. Department of Defense. (2005). *Defense Language Transformation Roadmap*. Retrieved from <http://www.defense.gov/news/mar2005/d20050330roadmap.pdf>
- U.S. Department of Defense. (2010). *Quadrennial defense review*. Retrieved from http://www.defense.gov/QDR/images/QDR_as_of_12Feb10_1000.pdf
- U.S. Government Accountability Office. (2011). *Language and culture training: Opportunities exist to improve visibility and sustainment of knowledge and skills in Army and Marine Corps general purpose forces*. (GAO-12–50). Retrieved from <http://www.gao.gov/assets/590/585990.pdf>
- Wang, X., Day, E. A., Kowollik, V., Schuelke, M. J., & Hughes, M. G. (2013). Factors influencing knowledge and skill decay after training. In W. Arthur, Jr., E. A. Day, W. Bennett, Jr., & A. M. Portrey (Eds.), *Individual and team skill decay: The science and implications for practice* (pp. 68–116). New York, NY: Routledge.
- Wilson, M. A., Drewes, D. W., Cunningham, J. W., Sanders, M. G., Thompson, J. A., & Surface, E. A. (2001). *Modeling SF soldier performance* [Technical Report]. Raleigh, NC.
- Wilson, M. A., & Sanders, M. G. (2003, April). *Project A Team: US Army Special Forces Assessment and Selection*. Symposium at the 18th annual meeting of the Society of Industrial and Organizational Psychology, Orlando, FL.

Received August 20, 2014

Revision received November 25, 2014

Accepted November 26, 2014 ■