Technical Report 2023-002

Deliverable - Leveraging Existing Department of Defense Data Towards Optimized Individual and Team Performance in the Army

Draft Final report for Cooperative Agreement W911NF120164

Joel Thurston

Joanna Schroeder

Nathaniel Ratcliff

Vicki Lancaster

Kathryn Linehan

Aaron Schroeder

Joshua Goldstein

Stephanie Shipp (PI)

University of Virginia,
Biocomplexity Institute and Initiative,
Social and Decision Analytics Division

December 2023

U.S. Army Research Institute
for the Behavioral and Social Sciences

Department of the Army
Deputy Chief of Staff, G1

Authorized and approved:

####
Director

Research accomplished under contract
for the Department of the Army by

University of Virginia, Biocomplexity Institute & Initiative, Social and Decision Analytics Division

Technical Review by

Kelly S. Ervin, U.S. Army Research Institute

DISPOSITION

This (insert document type) has been submitted to the

Defense Information Technical Center (DTIC).

REPORT DOCUMENTATION PAGE		
1. REPORT DATE December 2023	2. REPORT TYPE Final	3. DATES COVERED (from. . . to) 14 Jun 2019 - 14 October 2023
4. TITLE AND SUBTITLE Leveraging Existing Department of Defense Data Towards Optimized Individual and Team Performance in the Army		5a. CONTRACT OR GRANT NUMBER W911NF1920164 5b. PROGRAM ELEMENT NUMBER 0602785A
6. AUTHOR(S) Joel Thurston, Joanna Schroeder, Nathaniel Ratcliff, Vicki Lancaster, Kathryn Linehan, Aaron Schroeder, Joshua Goldstein, Stephanie Shipp (PI)		5c. PROJECT NUMBER 790 5d. TASK NUMBER 5e. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of Virginia, Biocomplexity Institute and Initiative, Social and Decision Analytics Division (SDAD) 1100 Wilson Blvd, Suite 2910, Arlington, VA 22209		8. PERFORMING ORGANIZATION REPORT
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral & Social Sciences 6000 6th Street (Bldg. 1464 / Mail Stop 5610) Fort Belvoir, VA 22060-5610		10. MONITOR ACRONYM ARI 11. MONITOR REPORT NUMBER
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.		
13. SUPPLEMENTARY NOTES ARI Research POC: Dr. Kelly Ervin		

14. ABSTRACT

The U.S. Army, the nation's largest military organization, seeks a comprehensive understanding of how the physical, cognitive, and social traits of Soldiers shape their ability to navigate the challenges presented by a persistently conflicted world. This research endeavors to pinpoint specific social characteristics of Soldiers, defining and quantifying them through the analysis of administrative data collected by the Army and now available to researchers in the Army's Person-Event Data Environment (PDE). The overarching goal is to evaluate the influence of these characteristics on both individual and unit performance. Drawing from the Army's extensive use of administrative and survey data, this paper provides a historical overview of individual performance assessment within the military. Additionally, it offers a comprehensive survey of the existing performance literature, culminating in the development of a conceptual framework. This framework serves as a foundation for subsequent quantitative and qualitative analyses, guiding the development of statistical models of Soldier and unit performance. The research identifies predictors and outcomes related to performance, concluding with a mixed model analysis that employs qualitative and natural language processing techniques. Findings reveal a consistency in the Army's descriptions of performance over time, underscoring persistent similarities in key social characteristics (e.g., a willingness to serve), despite an ever evolving set of specific task requirements. This study underscores the potential use of administrative data to measure performance and effectively connect it to meaningful outcomes, thereby contributing valuable insights to our broader understanding of performance and emphasizing the crucial role of social components in predicting performance outcomes.

15. SUBJECT TERMS

Soldier performance, unit performance, Global Assessment Tool (GAT), Person-Event Data Environment (PDE) benchmarking, social constructs, statistical models of performance, qualitative analysis.

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON
16. REPORT	17. ABSTRACT	18. THIS PAGE			
Unclassified	Unclassified	Unclassified	Unlimited	130	Kelly S. Ervin 703-545-2401

Technical Report ####

Deliverable - Leveraging Existing Department of Defense Data Towards Optimized Individual and Team Performance in the Army

Draft Final report for Cooperative Agreement W911NF120164

Joel Thurston

Joanna Schroeder

Nathaniel Ratcliff

Vicki Lancaster

Kathryn Linehan

Aaron Schroeder

Joshua Goldstein

Stephanie Shipp (PI)

University of Virginia,
Biocomplexity Institute and Initiative,
Social and Decision Analytics Division

December 2023

AUTHORS' NOTE

The views expressed in this article are those of the authors and do not reflect the official policy or analysis position of the Department of the Army, DOD, or the U.S. Government. All data, metadata, and complete scripts are made available with proper research approvals through the Person-Event Data Environment (PDE) managed by the Research Facilitation Laboratory (RFL) of the Army Analytics Group (AAG) and are cleared for public distribution per Army Directive 2021-18 (Use of People Analytics Data/Data Omnibus). Please send correspondence to Joel Thurston at jt9sz@virginia.edu.

LEVERAGING EXISTING DEPARTMENT OF DEFENSE DATA TOWARDS OPTIMIZED INDIVIDUAL AND TEAM PERFORMANCE IN THE ARMY

CONTENTS

ABSTRACT

The U.S. Army, the nation's largest military organization, seeks a comprehensive understanding of how the physical, cognitive, and social traits of Soldiers shape their ability to navigate the challenges presented by a persistently conflicted world. This research endeavors to pinpoint specific social characteristics of Soldiers, defining and quantifying them through the analysis of administrative data collected by the Army and now available to researchers in the Army's Person-Event Data Environment (PDE). The overarching goal is to evaluate the influence of these characteristics on both individual and unit performance.

Drawing from the Army's extensive use of administrative and survey data, this paper provides a historical overview of individual performance assessment within the military. Additionally, it offers a comprehensive survey of the existing performance literature, culminating in the development of a conceptual framework. This framework serves as a foundation for subsequent quantitative and qualitative analyses, guiding the development of statistical models of Soldier and unit performance. The research identifies predictors and outcomes related to performance, concluding with a mixed model analysis that employs qualitative and natural language processing techniques. Findings reveal a consistency in the Army's descriptions of performance over time, underscoring persistent similarities in key social characteristics (e.g., a willingness to serve), despite an ever evolving set of specific task requirements. This study underscores the potential use of administrative data to measure performance and effectively connect it to meaningful outcomes, thereby contributing valuable insights to our broader understanding of performance and emphasizing the crucial role of social components in predicting performance outcomes.

Keywords: Soldier performance, unit performance, Global Assessment Tool (GAT), Person-Event Data Environment (PDE) benchmarking, social constructs, statistical models of performance, qualitative analysis.

PART 1: INTRODUCTION

Goals of this Project

This cooperative research project (cooperative agreement #W911NF-19-2-0164: "Leveraging Existing DOD Data Towards Optimized Individual and Team Performance in the Army") between the University of Virginia Biocomplexity Institute's Social and Decision Analytics Division (SDAD) and the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) was undertaken to investigate how organizations – specifically the U.S. Army – could derive insights from large and disparate data sources to inform talent management decisions and strategic human resource planning. The work was motivated by a growing recognition of the untapped strategic value that rests in the stockpiles of administrative data held by the federal government (National Academies of Sciences, Engineering, and Medicine, 2017, 2019) and by broad trends across the Department of Defense (DOD) and other government agencies to adopt evidence-based policy-making (see H.R.4174, 2019; Department of the Army, 2018). This project explored the feasibility of using "Big Data" analytics as a means of predicting individual and team performance in a complex organizational setting.

The research leveraged the SDAD Data Science Framework (see Part 2: Applying the Data Science Framework), guided by a conceptual framework developed in conjunction with ARI (see Part 1: Conceptual and Methodological Profiling), to investigate, integrate, and wrangle disparate data sources. Such an undertaking proved to be a complicated technical endeavor. Once successfully integrated, descriptive analytics were used to provide a more cohesive assessment of Soldier performance as a function of individual, unit, and environmental contexts focusing on three main areas of interest:

1. What DOD and U.S. Army data can be used to create constructs for individual and team performance?
2. What DOD data contribute to, or predict unobservable constructs (e.g., team cohesion, collective identity, and commitment to organizational values)?
3. What DOD data best predict individual and team performance?

This research advanced the theory and science of workforce planning, as well as individual and team performance. The knowledge gained is valuable to the U.S. Army, and other large organizations, in addressing these questions, as well as informing the larger question of how data analytics can and should best be integrated into human capital management policies and programs that address individual and team performance.

A History of Performance Studies in the Army

Our most current understanding of performance is due in no small part to contributions from the U.S. Army. In August 1917, Secretary of War Newton Baker established the Committee on Classification of Personnel in the Army comprised of ten psychologists. The committee arguably developed the first rigorous personality tests to be used to predict future performance. Although ultimately not widely deployed, the tests were intended to gauge Soldiers' susceptibility to PTSD, what was then known as "shell shock," while serving in the first World War (Thulin, 2019). The Army conducted additional testing during World War I to select and classify both literate (Alpha Test) and illiterate recruits (Beta Test) (ARI, 2015).

This evaluative screening research would continue into WWII, following which much of the Army's interest would turn to personnel testing and test development. The Army's interest expanded into other realms of behavioral science research – like human engineering and physiological psychology – while also incorporating new techniques in testing and psychometrics. Research on noncognitive predictors began with the work of Borman who developed constructs for motivation, moral, and job satisfaction (Borman et al., 1975). But it was not until the 1980's that a long-term commitment (more than 10-years) was made to a holistic assessment of the Soldier that included noncognitive predictors to identify human differences with the congressionally mandated research program, Project A. (Campbell & Knapp, 2001).

From ASVAB to Project A

A major assessment tool still in use today, the Armed Services Vocational Aptitude Battery (ASVAB), was introduced in 1968 as part of the Student Testing Program. ASVAB scoring is based on Item Response Theory (IRT). IRT enables test questions and examinee abilities to be placed on the same scale, thereby allowing tests to be tailored to the specific ability level of each examinee and scores to be expressed on the same scale regardless of the combination of items that are taken. The IRT model underlying ASVAB scoring is a three-parameter logistic (3PL) model. The 3PL model represents the probability that an examinee at a given level of ability will respond correctly to an individual item with given characteristics. The item characteristics represented in the 3PL model are difficulty, discrimination (i.e., how well the item discriminates among examinees of differing levels of ability), and guessing (i.e., the likelihood that a very low ability examinee would respond correctly simply by guessing).

The ASVAB continues in use, periodically undergoing review and adjustment. Its prescores are primarily to determine (1) enlistment eligibility, (2) assign applicants to military jobs, and (3) aid students in career exploration. The ASVAB is also used to calculate a score on the Armed Forces Qualification Test (AFQT) from four ASVAB subtests: (1) arithmetic reasoning, (2) mathematics knowledge, (3), paragraph comprehension, and (4) word knowledge. Military services, such as the Army, use enlistees scores on the AFQT broken out into percentiles, to determine the level of high-quality recruits that are accessing.

In 1980 the ASVAB was standardized incorrectly resulting in 50% of the recruits being drawn from the bottom 30%. This led Congress to question the predictive value of the ASVAB in selecting high performing recruits (Fischl et al., 1980) and set in motion the validation of the ASVAB against measures of job performance across all military branches. The Army Selection and Classification Project or Project A was the Army's contribution. One of the goals of Project A – specifically Task 4 – was to construct Army-wide performance measures, that is, a measure of a Soldier's overall performance that was not tied to the various tasks required for a particular military occupation specialty (MOS).

Until Project A, research on predictor-criterion relationships was limited due to the scarcity of criterion measures such as adequate employee records (Pulakos et al., 1988). Predictor-criterion research relied on the "best" available criterion measures such as appraisal ratings rather than a more systematic approach that quantified the constructs of a well-thought-out and researched conceptual model, which could then be validated using statistical methods.

Identifying administrative in-service variables that could be used to construct an Army-wide performance measure was an integral part of Project A Task 4. In-service variables are those obtained after Soldiers access into the Army. For Task 4, the variables were limited to those variables that researchers (1) believed had potential to predict how a Soldier would perform in the future and (2) were contained in administrative data sources such as a Soldier's Enlisted Master File, Official Military Personnel File, and Military Personnel Records Jacket (MPRJ). The information that the project collected covered a Soldier's background data, enlistment conditions, progress in the Army, indicators of attrition, and commendatory and disciplinary actions.

In addition to issues inherent in evaluating these variables for fitness-for-use in constructing performance measures, the 1980's logistics of accessing the data sources and their timeliness were problematic. The Official Military Personnel File was maintained on microfiche located at the Enlisted Records and Evaluation Center (EREC) in Fort Benjamin, Indiana. Any additions or corrections to a Soldier's file had to be forwarded to EREC. Accessing the EREC data required visiting the facility. The other two data sources were maintained in individual personnel folders and required site visits to a Soldier's post. Updates and corrections to the MPRJ were made at the time of the action, but the Enlisted Master File only contained data from the preceding year. In the end, even though variable selection was ideally informed by theory, it was ultimately determined by practical considerations such as the timeliness of the data and the obstacles to accessing the information. These constraints led to data solely being drawn from the MPRJ.

Follow on research by Borman et al. (1987) took a different approach to constructing Army-wide job performance criteria. The goal of these researchers was not to assess performance using existing administrative variables but to develop constructs that described effective and ineffective first-term Soldier behavior independent of task performance. Like their precursors in Project A, they were guided by a broad conceptualization of job performance and defined a set of criterion behaviors that included elements of soldier effectiveness not directly related to task performance. Being a good soldier from the Army's perspective means more than just performing the job in a technically proficient manner.

Pulakos et al. (1988) used a multidimensional framework to construct validation strategies to explore the associations of cognitive ability and personality constructs (predictors) with different criteria. The purpose of their research was to demonstrate that multiple criterion factors can be identified in performance rating data. Individual-difference predictor measures were hypothesized to have different magnitudes of association with different criteria and ability to reveal predictable patterns. By examining these patterns, the resulting predictor-criterion correlations increase the scientific understanding of what both the predictors and criteria are measuring. Pulakos et al. (1988) demonstrated this by utilizing administrative data (as predictor variables) and collecting rating data for use as criterion variables.

Other Performance-Related Work in the U.S. Army

Select21 (2003-2008). In the early 2000s, the Army began the Future Force initiative to address new military challenges of the 21st century. Brigade Combat Teams were reorganized, which required personnel that possessed different combinations of knowledge, skills, and attributes (KSAs). The reorganization emphasized having the right Soldier in the right job. The goal of the Select21 program was to develop and validate new performance predictors for an entry-level selection and classification system adapted to the demands of the Future Forces 21st century initiative. The new measurement instruments were intended to supplement the ASVAB (what Soldiers "can-do") by predicting Soldiers "will-do" aspects of performance. The new predictors measure cognitive ability, temperament, psychomotor skills, values, expectations, and experience.

NCO21 (2004-2007). The NC021 research program was undertaken to help the Army plan for the impact of future demands on the noncommissioned officer (NCO) corps. The performance requirements and associated KSAs expected of future successful NCOs were used as a basis for developing tools that could be incorporated into an NCO performance management system, again geared towards demands of new 21st-century jobs.

Comprehensive Soldier Fitness Program (CSF) (2008-ongoing). The CSF program was established to address the challenges Soldier face due to multiple deployments required by persistent, prolonged conflicts across multiple geographic areas. Aimed at preventative measures for the Soldier, their families, and Army civilians, the program (renamed the Comprehensive Soldier and Family Fitness or CSF2 in 2012) is designed to build resilience and enhance performance through three main components: (1) self-development, (2) training, and (3) metrics & evaluation.

The CSF program is designed to increase psychological strength and productive performance and to reduce the incidence of maladaptive responses in the Army. There are four program elements:

(a) the assessment of emotional, social, family, and spiritual fitness using the Azimuth Check (formerly known as the Global Assessment Tool or GAT, see Table 14); (b) individualized learning modules to improve fitness in these domains; (c) formal resilience training; and (d) training of Army master resilience trainers (MRTs) to instill better thinking skills and resilience in their subordinates. CSF is proactive; rather than waiting to see who has a negative outcome following stress, it provides ways of improving resilience for all members of the Army. CSF aims to shift the current focus of how the Army and Soldiers respond to trauma and adversity—ranging from stress-related disorders to ordinary resilience—toward one of personal growth.

A key component of the CSF, the Azimuth Check/GAT is a survey for confidentially assessing psychological health based on five dimensions of fitness: emotional, social, spiritual, family, and physical. Active-Duty Soldiers are expected to complete the survey annually. It is optional for Department of the Army Civilians and Soldiers' family members. The Azimuth Check/GAT is a self-administered and self-evaluated mental health diagnostic tool. Built from existing, validated measures, the tool is intended to identify changes over time in resilience and coping skills, however, research points to the assessed characteristics largely remaining stable across time at the both the population and individual level (Ratcliff et al., 2022).

Tailored Adaptive Personality Assessment System (TAPAS) (2009-ongoing), TAPAS is a personality measurement tool developed by Drasgow Consulting Group under the Army's Small Business Innovation Research Program. The system builds on the foundational work of the Assessment of Individual Motivation (AIM; White & Young, 1998) by measuring narrow personality constructs (i.e., facets) that are known to predict outcomes in work settings. Utilizing computer adaptive features to assure each test is unique and results in more accurate individual assessments, the TAPAS uses a force-choice test method to reduce faking and measures 25 personality facets: can-do, will-do, achievement, adjustment, cooperation, dominance, even temper, excitement/attention seeking, intelligence efficiency, order, physical motivation, self-control, sociability, tolerance, selflessness/generosity, optimism/wellbeing. It is administered to enlistees at the Military Entrance Processing Stations (MEPS) along with ASVAB. Following in the tradition of the Select21 program, the TAPAS is designed to capture Soldiers' "will-do" spirit to predict future performance that more traditional cognitive- and vocational-focused assessments like the ASVAB might otherwise overlook.

Today, agencies such as ARI are charged with developing new means of improving and enhancing the Soldier lifecycle by bringing insight from the behavioral and social science into human resource policy decisions, while also exploring new domain that have the potential to affect the Army. Meanwhile, a focus on performance measurement, personnel assessment, and evaluating Soldier's psychosocial traits have proliferated across the service in a vast array of different programs.

Conceptualizing and Measuring Performance

Studying performance in the Army is difficult due in large part to the lack of a singular, universally accepted definition of what constitutes "performance" across the entire service. Various definitions or uses of the term appear in different doctrines and policies. Here are a few examples of these definitions or uses of the term "performance" from several DOD U.S. Army publications:

- Army Doctrine Publication (ADP) 6-22 - Army Leadership and the Profession: Performance is the execution of assigned tasks, skills, and behaviors to accomplish mission requirements. It encompasses various aspects such as physical, mental, and emotional capabilities, along with adherence to ethical standards and values.
- U.S. Army Training and Doctrine Command (TRADOC) Pamphlet 525-8-2 - The U.S. Army Learning Concept for 2015: This document defines performance in the context of individual and collective tasks, outlining it as the execution of specific duties, skills, and behaviors necessary to achieve desired outcomes.
- Army Doctrine Reference Publication (ADRP 7-0, Aug 2012) points to level of individualization (i.e., individual trainer) when it comes to defining performance: "Commanders and other leaders also use rehearsals to…understand how trainers intend to evaluate the performance of individuals and organizations and whether they understand how to conduct effective after-action reviews." (pp. 3-10)
- Joint Publication (JP 3-0, Jan 2017) defines Measures of Performance (MOPs) as a function of Commander-Centric Leadership; highlights role of CO in determining performance in context of accomplishing tasks to a standard.

One theme running throughout these definitions is that performance is tied to the achievement of outcomes and task accomplishment. Performance is functionally achieving a specific end state (i.e., meeting required conditions to complete objectives). While useful for triangulating on conceptual agreement, this by necessity makes the act of performing context dependent, which limits the ability to objectively assess performance across different contexts.

Delving further, we see that task accomplishment within these contexts are often treated as a binary variable. For example, ADRP 7-0 (Aug 2012), in discussing Training and Evaluation Outlines (T&EO) states: "All training must be evaluated…[T&EO]… provides information on collective training objectives, related individual training objectives...this document provides the task title, task description, the recommended conditions to use in training, the standard to be met and the task steps and *performance measures* to attain a 'GO/NO-GO' for each step" (pp. 3-12). We see similar binaries on Mission Essential Tasks (METs) rated as Go/No-Go or Weapons System Proficiency as Qualified/Not Qualified. When not measured as a binary, performance is still treated categorically. Scoring on the Objective Task Matrix (OBJ-T) uses three categories of Trained, Practiced, and Untrained. Even in cases where there is implicit acknowledgement that performance falls on a spectrum, the spectrum is only important to the degree that a minimum acceptable level is established. In the ADRP 7-0 (Aug 2012), for example: "Each individual and collective task has standards of performance. A standard is the accepted proficiency level required to accomplish a task" (pp. 2-2). This determines a minimum threshold for clearance but does not capture the degree to which higher levels of performance may be achieved.

This is not to dispute the usefulness of performance as used in a military context. These definitions of performance are necessary to facilitate the Army's mission. Training to a standard ensures that all Soldiers are minimally competent to perform the required tasks of being a Soldier or operating in a specific MOS. However, a binary or categorical definition of performance does not align with the more nuanced view that has emerged from academia beginning in the 1980's. This is not to say that there is complete agreement among all academic definitions either (there is not), simply that most modern theories of performance allow for a spectrum of responses. This is largely due to the different end goals of the users. Researchers seek to explore every facet of a phenomena to accurately measure it and map its component parts in as much detail as possible, while the Army must fight and win the nation's wars. Performance is therefore treated or defined differently in these contexts to facilitate their desired end states.

Appropriately, this focus on results highlights another difference in the way that these groups (broadly defined) consider the concept of performance. Due to the lack of lack of conceptual uniformity across the use of performance in Army writing and its use in different levels of doctrine (i.e., strategic, operational, and training), performance or the act of performing is often defined as a function achieving a particular outcome. Consider Measures of effectiveness (MOE) and Measures of performance (MOP) from Joint Publication (JP) 3-0:

- MOE: "A criterion used to assess changes in system behavior, capability, or operational environment that is tied to measuring the attainment of an end state, achievement of an objective, or creation of an effect." (JP 3-0)
- MOP: "A criterion used to assess friendly actions that is tied to measuring task accomplishment. (JP 3-0)" is a task-specific, not holistic, assessment.

Achieving a specific outcome, however, is not something most academic definitions would consider performance (Campbell et al.; Koopmans et al.).

Academia, on the other hand, posits a much more complicated structure, often involving different types of performance and typically (and explicitly) divorcing the act of performing from the potential outcomes of performance. This is true even of approaches that can directly trace their lineage to foundations in Army work such as Project A. SDAD performed an extensive literature review across the performance literature to gain a detailed understanding of how different disciplines conceptualize and measure performance. We focused predominantly on those researchers who had built performance taxonomies (Campbell et al., 1993, Campbell et al., 2015, Koopmans et al., 2011, Sackett et al., 2017), because we could leverage the work already spent synthesizing multiple perspectives while building our own comprehensive model of performance to date. In doing so, we could assess the relative value of each perspective (e.g., was it focused on a particular type of performance, was it focused solely on how individuals or groups performed) and how best to incorporate it into our own framework. In doing so, we identified the most comprehensive models of performance to date, which served as the foundations for our conceptual framework.

Campbell et al. (1993 & 2012)

In 1993, Campbell and his colleagues published work extending the Project A model of performance to make it more appropriate for non-military jobs. They identified a latent structure of performance consisting of an eight-factor model:

1. Job specific task proficiency
2. Non-job specific task proficiency
3. Written and oral communication task proficiency
4. Demonstrating effort
5. Maintaining personal discipline
6. Facilitating peer and team performance
7. Supervision/leadership
8. Management/administration

Importantly, Campbell et al. (1993) declared that performance was synonymous with behavior. Performing is something that people do and, therefore, can be observed. Consequently, as an observable phenomenon, this also meant that performance could be scaled (i.e., measured) and that each person's individual level of proficiency determined (Campbell et al., 1993). However, this definition of performance only included behaviors relevant to an organization's goals.

In 2012, Campbell revised the conceptualization to represent performance as a consensus latent structure that is described as concretely as possible. That is, the intent was to use as few difficult-to-define abstractions as possible, even though it makes things sound less exciting. The eight factors were also updated:

1. Technical performance
2. Communication
3. Initiative, persistence, and effort
4. Counterproductive work behavior
5. Supervisory, managerial, executive leadership
6. Hierarchical management performance
7. Peer/team member leadership performance
8. Peer/team member management performance

Koopmans et al. (2011)

Koopmans et al. (2011), drawing in part on Campbell et al. (1993), completed a review of performance literature in occupational health, psychology, and management to identify conceptual frameworks of individual work performance to formulate a heuristic conceptual framework. They propose a framework of individual work performance in which overall work performance consists of four dimensions: task performance, contextual performance, adaptive performance, and counterproductive work behavior. In this model performance was the result of trait, state, and situational factors – such as a person's cognitive ability, attitude, and organizational reward structure – mediated by direct determinants such as skills or choice behavior. Like Campbell and colleagues before them, Koopmans et al.'s model draws a distinction between performance (actions within the full control of the individual) and outcomes, which could be influenced by additional external factors (e.g., sales dropping due to a downturn in economic conditions).

Based on this work, in conjunction with discussions with subject matter experts from ARI, we identified the following definition of performance to use in building a conceptual framework to understand performance in the Army.

> Work performance is an *abstract, latent construct* that cannot be pointed to or measured directly. It is made up of multiple components or dimensions. These dimensions… are made up of indicators that can be measured directly. To conceptualize and operationalize individual work performance, we should explicate the construct domain of work performance and identify its dimensions and indicators. Whereas the dimensions may generalize across jobs, the exact indicators can differ between jobs. (Koopmans et al., 2011)

Analyzing performance items with natural language processing (NLP)

In addition to our literature review, and as a precursor to what would become Analysis 2, we tested the usefulness of natural language processing (NLP) to augment the work of human reviewers and coders. Our goals were twofold: (1) demonstrate that NLP techniques were effective in identifying latent concepts from limited text input and (2) if successful, compare the concepts identified against the elements of performance outlined by Campbell et al. (1993) and Koopmans et al. (2011).

Working with a team of undergraduate and graduate student researchers during our Data Science for the Public Good Young Scholars program, we conducted an extensive review of published performance measurement scales. The team identified 63 scales drawn primarily from academic publications (approximately 75%) with an additional 15% coming from non-DOD government sources, and the remainder from industry, military, and other sources (Data Science for the Public

Good, 2020). Nine of the scales were developed pre-1980, but the majority represented work done in the last 40 years[1].

Once the corpus of scales was assembled, we tested our ability to extract individual items from within the scales and identify thematic connections of item clusters based on specific concepts (e.g., items evaluating cognitive ability vs. motivation). We utilized two NLP topic modeling approaches in identifying thematic clusters of items across the scales: Latent Dirichlet Allocation (LDA) and Biterm Topic Modeling (BTM).

LDA is a probabilistic topic modeling process which considers each document (i.e., scale item) as a distribution of topics and each topic as a distribution of words (Blei et al., 2003). Using Dirichlet probability distribution, the model draws out the implicit themes across a set of documents. We implemented LDA using the STM package in R. LDA works best with larger data sets, which was a challenge given the relatively limited corpus of items we had available. More information about this technique can be found in Part 3, Analysis 4 of this report.

We compensated for this challenge by not removing typical "stop words" prior to analysis. Stop words are extremely common words which help to form the structure of sentences but do not inform the content. Leaving the stop words in the dataset also contributed to the analysis in a second way. By including pronouns such as "I" and "you," we preserved valuable conceptual information about the scale items, such as the level of scales the scales were intended to assess (i.e., individuals vs. groups). Acknowledging that some common words provided a less beneficial tradeoff of conceptual value (e.g., to, the, in), we constructed a custom stop word list to remove by examining the most frequent words in our data set. We also removed words that referred to specific jobs (e.g., nurse) and locations (e.g. Britain, Newfoundland).

The LDA outputs revealed several latent themes in the job performance scales. We identified themes related to ideal job characteristics, feelings about supervisors and organizational climate, satisfaction, health and wellness, families, and working in groups. Notably, many of these themes aligned with the non-task performance types identified by researchers (e.g., counterproductive performance, contextual performance). Other themes touched on different aspects of the social dimensions that would eventually be folded into our conceptual framework in other areas (e.g., as predictors and outcomes).

We performed a similar analysis using the BTM approach to topic modeling. BTM is a method for detecting topics occurring in short texts (Yan et al., 2013). As we saw with LDA, a challenge in working with traditional topic modeling techniques is that very short documents may not have enough data to provide interpretable results. Our set of documents were individual scale items, essentially the smallest type of document imaginable in that many "documents" were only one sentence long. BTM addresses this problem by analyzing the entire corpus of documents together,

[1] Four of the scales could not be traced to an original publication date.

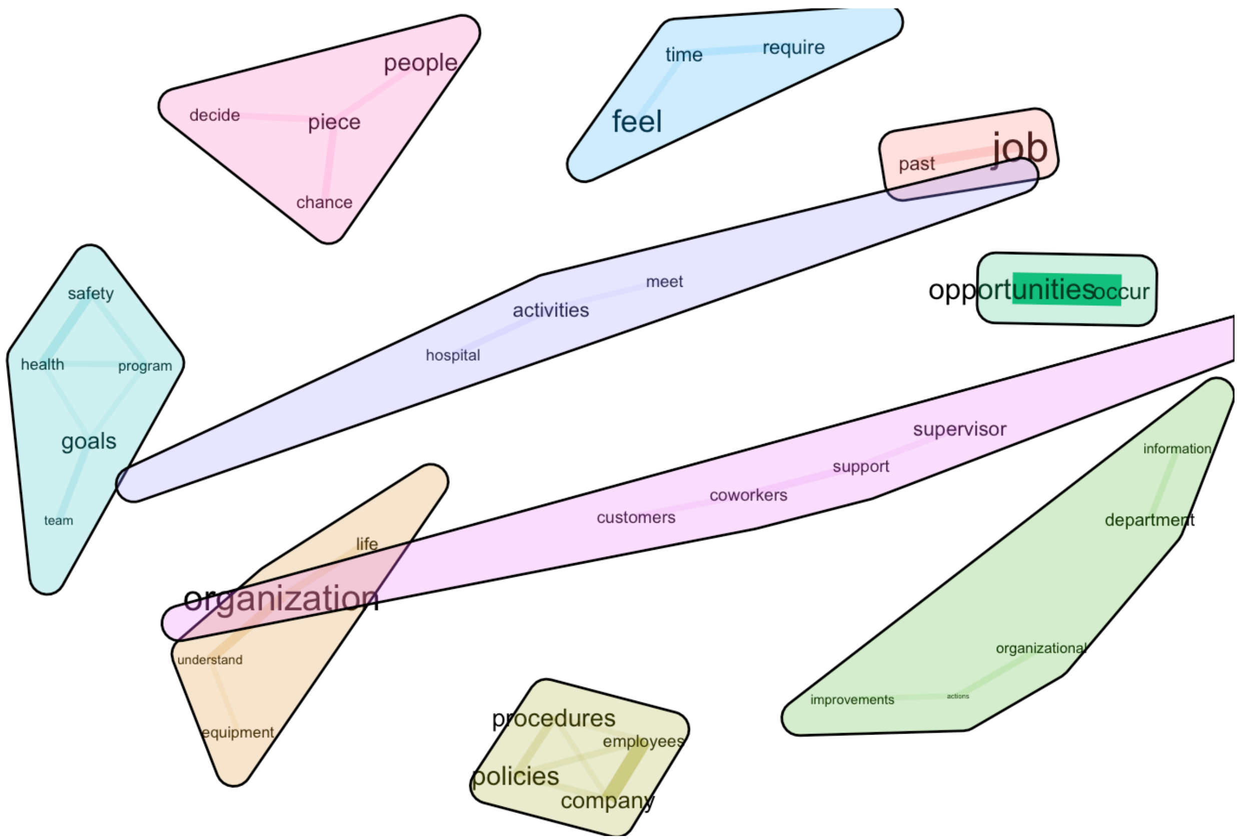
focusing not on individual documents but on collections of "biterms." A biterm is an unordered co-occurrence of words.

As with LDA, BTM allows the user to specify the number of topics to cluster the corpus around. We chose 10 items as a test case, averaging the number of components we observed across the most complex models of performance which typically specified eight to 12 elements of performance. Our results provided both a list of the biterms most closely associated with each topic, as well as a visual representation of the topics and their most frequently occurring words. In Exhibit 1 we see all 10 topics, including ones related to:

- Organizational policies (diamond, lower center)
    - Other words observed: Adapt, fair, job, performing, priorities
- Temporal elements of performance (triangle, upper right)
    - Other words observed: Job, overtime, task
- Relationships across organizations (flat triangle, lower middle)
    - Other words observed: Job, listen, opportunities, perform, solving, talk
- Opportunities (rectangle, right)
    - Other words observed: Abilities, advancement, duties, environment, interact
- Health and safety (diamond, left)

These results provided additional insight into our conceptual framework, affirming the notion that performance, even when assessed at the individual level, is interconnected with the place a person occupies within a larger context (e.g., their team, their organization). It also confirmed an interconnectedness between types of performance (e.g., contextual, adaptive, counterproductive) and the interplay between characteristics typically viewed as inherent to an individual (e.g., personality traits) and exogenous factors, as well as choice behavior.

Exhibit 1.
Results of the Biterm Topic Model Analysis



Our Conceptual Framework

Distilling extensive literature review SDAD performed on military and non-military performance research, we developed a hierarchal performance framework (see Exhibit 2). The framework builds on existing performance taxonomies (e.g., Campbell et al., 1993; Campbell & Wiernik, 2015; Koopmans et al., 2011) and operationalizes them into a framework that is more relevant to Soldier performance. It is grounded in the definition of performance as a behavior or action that people engage in to further the goals of the organization that can be quantified or scaled (measured) in terms of each individual's proficiency (i.e., level of contribution) or as a collective action at the group level. The individual Soldier is modeled as a combination of characteristics. These characteristics include trait and state factors.

- Trait factors are relatively stable characteristics such as cognitive skills, personality type, and a Soldier's physical condition.
- State factors are characteristics that can or do change more easily across time such as knowledge, skills, attitudes, perceptions, job satisfaction, and motivation.

These characteristics serve as direct determinants of performance in that they are qualities that have an impact on performance. Our model recognizes that there are other factors, potentially not captured by the data we have available, that will undoubtedly also influence Soldier performance. Additionally, the model recognizes that many of these characteristics likely interact with one another in some form – essentially functioning as mediators and moderators in a modeling context.

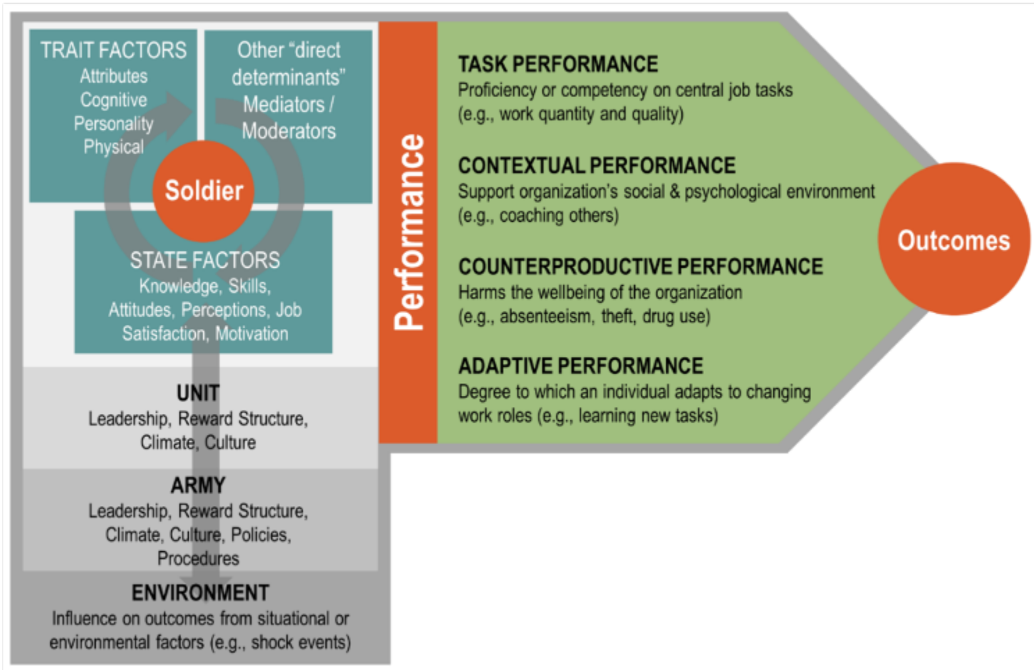
The model is hierarchical because it recognizes that Soldiers are members of Army units. Units possess characteristics all their own, which will influence or interact with an individual's characteristics to further influence Soldier performance. Examples of a unit's characteristics include culture, climate, leadership, and reward structure (i.e., behaviors that are recognized and valued) within a unit. Units are further nested within the overall Army. Each subsequent level of the Army hierarchy (including the organization in its totality) possesses characteristics that can be defined in the same way but on a larger scale. Finally, the model acknowledges that there are external environmental factors that can impact performance outcomes, especially occurrences such as represent shock events (e.g., 9/11 terror attack), changes in the economy, and other exogenous changes that the Soldier, units, and Army cannot control.

These layered factors come together to manifest as Solider and unit performance, which we categorize into four different types as described by Koopmans et al. (2011) and listed below.

- Task performance is proficiency or competency on central job tasks.
- Contextual performance supports the organization's social and psychological environment.
- Counterproductive performance harms the well-being of the organization.
- Adaptive performance is the degree to which an individual adapts to changing work roles.

The intersection of these performance behaviors determines the overall performance along a spectrum of outcomes. However, as noted in our guiding definition of performance, we maintain outcomes are separate from the act of performing, because outcomes are subject to influence outside the control of the agent (whether that agent is a single Soldier or a unit) alone.

Exhibit 2.
Conceptual Performance Framework



PART 2: LEVERAGING ADMINISTRATIVE DATA TO STUDY PERFORMANCE

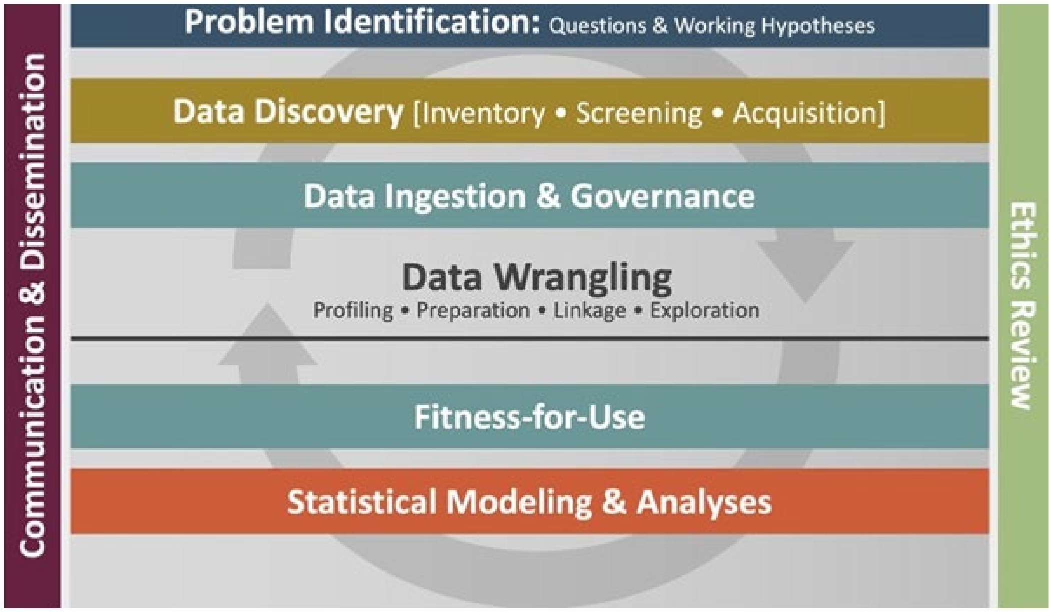
Studying performance in an Army context is conceptually difficult, as has been noted, but it also presents practical challenges. Rumsey (2012) outlines several of these noted by the Army directly, including recognition that (1) no single measure can completely capture all the information about a Soldier's performance, (2) testing (even hands-on) does not capture information about a Soldier's day-to-day performance, (3) input from others (e.g., peers, supervisors) is needed to capture some elements of a Soldier's performance (e.g., proficiency, motivation), and (4) job-tests, while useful, are not always administered effectively. Enter data science and the science of all data.

Data science is a transformative way to find meaning in a complex world (Provost & Fawcett, 2013; van der Aalst, 2020). Data science is an evolving field that combines methodology focused disciplines (e.g., statistics, computer science), subject matter expertise (e.g., social, psychological, physical, geolocation), and multiple levels of analysis (e.g., individuals, groups, states; Garber, 2019; Wing et al., 2018). Although frequently described as 'big data,' data science involves the use of data of shapes, sizes, and types. Much of this data may be repurposed from other sources (e.g., administrative data), but it can also include non-traditional forms of secondary data as well (see Adjerid & Kelly, 2018; Keller et al., 2020; King et al., 2016), in addition to opportunity and process data and even direct data collection. The questions data science seeks to answer are complex and multifaceted. Therefore, the most comprehensive answers will involve a combination of all types of data.

Bringing together such disparate data sources requires an organizing framework – something to guide the processes by which data are discovered, accessed, ingested, repurposed, and statistically integrated– a *data science framework* (Keller et al., 2018; 2020; cf. discussion of the data life cycle in Berman et al., 2018). Through years of work with stakeholders and grounded in the need to solve practical problems, the Social and Decision Analytics Division at the UVA Biocomplexity Institute has developed an operational framework for practicing applied data science (see Exhibit 3). Our framework is a comprehensive process that underlies the foundation of our research method. The process is rigorous, flexible, and iterative. Each stage is informed by the insight of prior stage. Importantly, although we describe the framework in a linear fashion, it is actually iterative (denoted by the inclusion of the circular arrow), indicating that insights from "later" stages are incorporated back into our understanding as we return to "earlier" stages of the process as the research continues. Importantly for practitioners and researchers alike, applying the Data Science Framework ensures consistency and transparency to the process, which in turn ensures reproducibility of results.

Exhibit 3. Data Science Framework

UVA BI Social and Decision Analytics Division Data Science Framework



Our framework processes four unique characteristics not typically seen in these processes. First, we explicitly focus our approach on answering specific research questions. Second, data discovery is a primary activity. Third, we consider the role of data governance to be an essential part of the research process and one that that builds trust with all stakeholders involved through the use of data-sharing protocols. Finally, we incorporate ethics into all components of the framework. Although a staple of social and behavioral sciences for the past several decades, ethical considerations have been slow to take their proper place within some of the computational fields (Keller et al. 2020)

PDE Overview, Access, and Issues

Person-Event Data Environment (PDE)

The Army's Person-Event Data Environment (PDE) is a secure data enclave used to access data sources for this study. The PDE is a remote-access, virtual data enclave that houses unclassified, but sensitive, secondary data assets sourced from various DOD systems and organizations including data about psychological measures, performance indicators, medical information, and administrative personnel records across individual service members' careers (National Academies of Sciences, Engineering, and Medicine, 2017; Vie et al., 2015). The Army Analytics Group's (AAG) Research Facilitation Laboratory (RFL) administers the PDE. It provides access to researchers and institutions with support and approval by Army or DOD sponsors (Knapp et al., 2018a). The research project needs determine what data are made available to researchers in conjunction with negotiation between RFL and the data owners with proper study protocol approvals from an Institutional Review Board (IRB) and the Army Human Research Protections

Office (AHRPO). Personally Identifiable Information (PII) is protected in the PDE through a Rosetta process which assigns a unique project specific PDE-generated personal identifier (PID) to every individual represented by the data. The PID for a given individual remains the same across data sources for a project enabling the linkage of multiple data types but is different across research projects in the PDE to protect privacy.

Breaking new ground and establishing new processes

This project continued our trailblazing work in accessing and utilizing the PDE. As outlined in the final report for the ARI sponsored project "Using Administrative and External Data Sources to Model First Term Attrition of Army Enlisted Soldiers" (W911NF-15-2-0133) there have been many technical hurdles to overcome in completing the research presented in this document. Access to data in the PDE continues to require a Common Access Card (CAC) with all work being done on the PDE server. For research staff that have previously had a security clearance, acquiring a CAC typically takes 60 to 75 days. For researchers with no prior security clearance, the average length of time required has been much higher (i.e., 180-240 days). Additionally, although only aggregate results are permitted to be downloaded for use in reports and publications, we have continued to experience delays in exporting these files from the PDE servers.

We continue to help drive change in AAG RFL's operating procedure. For example, we brought federal statistical data into the PDE, including the US Census Bureau's American Community Survey. We were one of the first projects to migrate a portion of our research to their cloud PDE computing environment (cPDE). However, in doing so, AAG RFL had to develop a means by which the results from these analyses could be imported back into the Analysts PDE (PDE-A) where most of our modeling occurs.

Conceptual and Methodological Profiling

Data science research involves ingesting and linking disparate sources of secondary data. While these sources can typically be cleaned and wrangled into a usable form for analysis, robust documentation on how the variables were created and the variables' intrinsic meaning is not always available or apparent. Raw numbers and text can only inform researchers about the state of being or what is about the data, not necessarily answering questions of meaning like the for what the data represent and how data were collected or recorded. In addition to profiling to assess the quality and format of data sources, it is essential to understand what concepts the variables represent and the methodology or process that produced the data (i.e., data provenance). Data provenance is a subject of increasing relevance to data science pipelines as a type of metadata that provides a contextual history of data and its relationship with data management systems (Doan et al., 2012; Glavic & Dittrich, 2007; Simmhan et al., 2005; Song et al., 2019). Over time, data can have a complex history involving numerous changes from its original source by being imported, transformed, or re-translated within and between data systems (Glavic & Dittrich, 2007). Along with other metadata, provenance provides the context to explain the origins of data which can build authenticity and trust in how to make sense of data and how it can be reused (Simmhan et al., 2005).

Importantly, contextual information is needed to guide how variables should be interpreted and used in subsequent analyses. Since data do not remember where they come from (Lord, 1953, p.21), data can be manipulated in any way that is mathematically feasible when conducting statistical analyses (e.g., addition, multiplication, regression) because these tests do not consider the objects or events to which the data refer. However, when it comes time for interpretation after an analysis, the question arises as to whether the results bear any meaningful relationship to the original objects or events being studied and thus, a conceptual/methodological issue arises rather than a statistical one (Howell, 2008, p. 21). Stated differently, results can be derived from a mathematically-sound statistical test, but this does not ensure that the methodology or conceptual meaning behind the test was sound or valid.

Without this key contextual knowledge, data can lack the context necessary to understand the best ways to use and analyze it, risking misinterpretation or misuse. In the absence of pre-existing metadata and provenance, important conceptual and contextual information needs to be derived. Therefore, in addition to profiling the quality and structure of the data sources, we performed conceptual and methodological profiling. Conceptual and methodological profiling provides a methodological framework for deriving this information and complements existing data profiling methods.

Conceptual profiling is a qualitative categorization process that involves identifying what constructs variables represent and what they measure. We define constructs as abstract concepts that are unobserved (i.e., not directly measurable or latent) and made concrete (indirectly) through operationalization of observed measures or indicators (Kline, 2011). The conceptual profiling categorization process included identifying the variable's construct (e.g., naming unobserved or latent variable the observed variable represents), level of analysis (e.g., individual-level, group-level), measure category (e.g., attribute, perceptual, behavioral), measure type (e.g., trait, state, outcome, process), and operational definition of the variable (i.e., how the variable is concretely defined). Conceptual profiling can be informed by past research about the variable or measure, the original intent of the author(s) who designed the measurement of the variable, and general personal expertise. In the cases where the variables were a part of a designed data collection (e.g., survey, experiment), there can be discrepancies between what the original research wanted the variables to represent conceptually and what they might better represent in the way they were measured.

Methodological profiling is a qualitative categorization process that involves identifying how the variables were created and how they were measured and/or recorded. The methodological categorization process can include identifying how the data were obtained (e.g., administrative, designed, opportunity), when the variable was collected (e.g., date or question position on a survey), from whom was the data collected (e.g., sample population), the item or question text (e.g., item stem and/or text), and the response type and values (e.g., Likert, free response, categorical). Together, conceptual and methodological profiling helped to identify variables of interest for further cleaning and modeling (see Exhibit 3).

Exhibit 4.

Example of Conceptual and Methodological Profiling of Variables

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	P	T	V	W
1	[PDE] ENT_NAME	[PDE] VAR_NAME	[PDE] VAR_BUSNAME	[PDE] VAR_DESCRIPTION	[PDE] VAR_USAGE	Table Name	Source	Table Type	HOC ID	Construct ID	Construct Referent	Construct Form	Construct Framework	Performance	Data Type	Operational Definition	Response Format	Response Values
2	MV_MASTER_AD	ACC_PGM_SRC_C	Accession Program	This is a DMDC derived	Applicable to	master	DMDC	administra		accession	individual	attribute	trait	na	administra	Reports information	categorical	1 = induction; 2 =
3	MV_MASTER_AD	ACC_SRC_IND_CD	Accession Source	(Indicator Code (non-		master	DMDC	administra		accession	individual	attribute	trait	na	administra	Reports information	categorical	empty
4	MV_MASTER_AD	ACS_SCRTY_CLRN	Access Security	The data is received daily	This data element	master	DMDC	administra		person	individual	attribute	trait	na	administra	Reports security	categorical	0 = no clearance; 1 =
5	MV_MASTER_AD	ADSVC_PE_DT	Active Duty Service	The date for which a DoD	Prior to October	master	DMDC	administra	event	event date	individual	attribute	trait	na	administra	Reported date of event.	free response	event date
6	MV_MASTER_AD	AD_IVL_RTNT_RS	Active Duty	The data is received daily	No Air Force	master	DMDC	administra		involuntary	individual	attribute	state	na	administra	Reported involuntary	categorical	1 = extension was
7	MV_MASTER_AD	AD_LRP_INCN_PG	Active Duty Loan	The data is received daily	A = Educational	master	DMDC	administra		person debt	individual	attribute	trait	na	administra	Reported debt or		1 = educational loan
8	MV_MASTER_AD	AD_LRP_INCN_PG	Active Duty Loan	The date on which a		master	DMDC	administra	event	event date	individual	attribute	trait	na	administra	Reported date of event.	free response	event date
9	MV_MASTER_AD	AD_LRP_INCN_PG	Active Duty Loan	This date indicates when		master	DMDC	administra	event	event date	individual	attribute	trait	na	administra	Reported date of event.	free response	event date
10	MV_MASTER_AD	AFMS_BASE_DT	Active Federal	The date for which DoD	Member Active	master	DMDC	administra	event	event date	individual	attribute	trait	na	administra	Reported date of event.	free response	event date
11	MV_MASTER_AD	AFMS_MN_QY	Active Federal	The numeric value of		master	DMDC	administra		length of	individual	attribute	state	na	administra	Reports the length of	free response	numerical
12	MV_MASTER_AD	AFMS_YR_QY	Active Federal	The data is derived from		master	DMDC	administra		length of	individual	attribute	state	na	administra	Reports the length of	free response	numerical
13	MV_MASTER_AD	AFM_OSVC_BASE	Active Federal	The date for which DoD	Applicable only to	master	DMDC	administra	event	event date	individual	attribute	trait	na	administra	Reported date of event.	free response	event date
14	MV_MASTER_AD	AFQT_CAT_CD	Armed Forces	The data is received daily	Only applicable to	master	DMDC	administra	cogniti	AFQT overal	individual	cognitive; index	trait	na	administra	Reported Armed Forces	categorical	1 = Category V--AFQT
15	MV_MASTER_AD	AFQT_PCTL_SCR_	Armed Forces	The data is received daily	Only applicable to	master	DMDC	administra	cogniti	AFQT overal	individual	cognitive; index	trait	na	administra	Reported Armed Forces	index	numerical
16	MV_MASTER_AD	AF_PAS_CD	Air Force Personnel	No definition provided by		master	DMDC	administra	uniden	unidentifiabl	na	na	na	na	administra	Reported	na	na
17	MV_MASTER_AD	AQD_2_CD	Additional	No definition provided by		master	DMDC	administra	uniden	unidentifiabl	na	na	na	na	administra	Reported	na	na
18	MV_MASTER_AD	AQD_3_CD	Additional	No definition provided by		master	DMDC	administra	uniden	unidentifiabl	na	na	na	na	administra	Reported	na	na
19	MV_MASTER_AD	AQD_4_CD	Additional	No definition provided by		master	DMDC	administra	uniden	unidentifiabl	na	na	na	na	administra	Reported	na	na
20	MV_MASTER_AD	ASG_BASE_FAC_I	Assigned Base Facility	(Base Identifier Code)	As a DMDC	master	DMDC	administra	geospa	base	unit	attribute	situational	na	administra	Reported base or post	categorical	codes:
21	MV_MASTER_AD	ASG_UIC_PDE	[PDE] Unit	The Servicemember's		master	DMDC	administra	organiz	unit	unit	attribute	state	na	administra	Unit Soldier was a	categorical	alphanumeric
22	MV_MASTER_AD	ASG_UNT_LOC_IS	Assigned Unit	(ISO Alpha 3 Country		master	DMDC	administra	geospa	unit location	unit	attribute	situational	na	administra	Reports location of unit	categorical	text

Note. PDE = Person-Event Data Environment, ENT = entity, VAR = variable, BUSNAME= business name, HOC ID = higher order construct identification, ID = identification.

These profiling processes can be performed in either order but are probably best done concurrently. Importantly, each categorization process is flexible and can be tailored to specific research needs by adding or subtracting the suggested qualitative taxonomies outlined below. While our research focuses on individuals and groups, this process could be used for other data domains such as financial (e.g., stocks), non-human (e.g., animal behavior), physical (e.g., climate measurements), or mechanical processes (e.g., machine functioning).

We conceptually and methodologically profiled over 3,500 variables. This process was performed on all data tables in the PDE provisioned to our project. Variables were categorized using various typologies (e.g., data type, construct identification, construct form) with different classification categories. Importantly, variables were categorized as to whether they met one of five performance dimensions (i.e., task, contextual, counterproductive, adaptive, and general) identified by Koopmans and colleagues (2011). Then, a group of four additional independent raters categorized a random subset of 156 variables which represented about 5% of the total corpus of variables. Lastly, measures of agreement were analyzed between the five raters (original rater plus four independent raters) revealing fair to excellent levels of agreement (Fleiss' Kappas = .41–.70). In cases where there was not a majority agreement on variables (i.e., 3 or more raters choosing the same category), points of disagreement were resolved through discussion.

One novel outcome to spring from this project has been our work more fully integrating techniques from the social and behavioral sciences (e.g., conceptual and methodological profiling) more fully into the Data Science Framework. This represents a more complete blending of two approaches often seen as distinct ways of understanding data but that, when used in combination, provides a more complete understanding of the value and limitations of a particular dataset to answer researchers' questions of interest (Ratcliff & Thurston, 2022).

Formation of Data Set

All data were accessed and analyzed in the Army's Person-Event Data Environment (PDE) using the statistical software R 3.6.1 (see the "Benchmarking Crosswalks of Army data from Multiple Sources" report for example code) (Ratcliff et al., unpublished). Data were taken or derived from different data sources in the PDE (see Exhibit 4). Individual-level data were linked between tables using a PID or Person Identifier, which is a de-identified marker for each unique person in the PDE that is also unique to each project within the PDE (i.e., the same individual in the PDE will have different PIDs for different projects). Once data had been ingested for our project, variables were profiled for their quality and structure (e.g., completeness, consistency, uniqueness, linkability). Data tables were also examined for their relative coverage over time; filing dates, report dates, and test dates were used to determine the earliest and latest dates represented by the data. This analysis was important for deriving the sample time frame (see Sample Selection Criteria below) because we needed to ensure that different pieces of data had adequate overlap to be used together in statistical modeling. For certain predictor variables (e.g., Rank Group), this meant using variables upon the Soldier's accession. For outcome variables (e.g., Award Count), values could be derived past the sample time frame because the outcomes represent the fullest accounting available for a Soldier's career performance.

Exhibit 5.

Data Tables Used from The Person-Event Environment (PDE)

Table Name	Source Name	PDE Table Name	Description	First Date	Last Date
Master	Active Duty Military Personnel Master	MV_MASTER_AD_ARMY_QTR_V3A	Master administrative records	2001-09-30	2019-12-31
MEPCOM 1	Military Entrance Processing Command	MEPCOM_USAREC_RA_ANALYST	Initial entry records	2000-10-01	2016-07-19
MEPCOM 2	Military Entrance Processing Command	MV_DMDC_MEPCOM_700_V2	Initial entry records	1999-09-30	2019-11-30
TAPAS	Tailor Adaptive Personnel Assessment	DMDC_TAPAS_201602	Personality test for placement upon entry	2010-03-01	2015-05-01
GAT 1.0	Global Assessment Tool (Active Duty Soldier)	GAT_SOLDIERS_V2	Psychosocial characteristics assessment	2009-05-05	2014-01-29
GAT 2.0	Global Assessment Tool (Active Duty Soldier)	GAT_SOLDIERS_20_V2	Psychosocial characteristics assessment	2013-09-09	2017-09-30
APFT	Army Physical Fitness Test	TA_DTMS_APFT	Physical fitness test scores	2001-01-21	2016-06-13
Height/Weight	Height & Weight	TA_DTMS_HT_WT	Height and Weight Test	2001-01-15	2016-06-13
Derogatory Statements	Interactive Personnel Elective Records Management System	TA_IPERMS_DEROG_V2	Negative papers and statements on record	2001-01-01	2018-06-16
Awards	Army Work Force Transaction File	MV_AWTF_AWARDS	Awards given records	2012-03-28	2018-12-31
Health 1	Medical Operational Data System	TA_PHA_OLDFORM_V1	Health records	1982-09-10	2017-04-01
Health 2	Medical Operational Data System	TA_PHA_NEWFORM_V25	Health records	2007-12-01	2017-03-20
Transaction	Active Duty Military Personnel Transaction	MV_TRANS_AD_ARMY_30_V3A	Entry and exit status within the Army	2001-09-01	2018-12-31

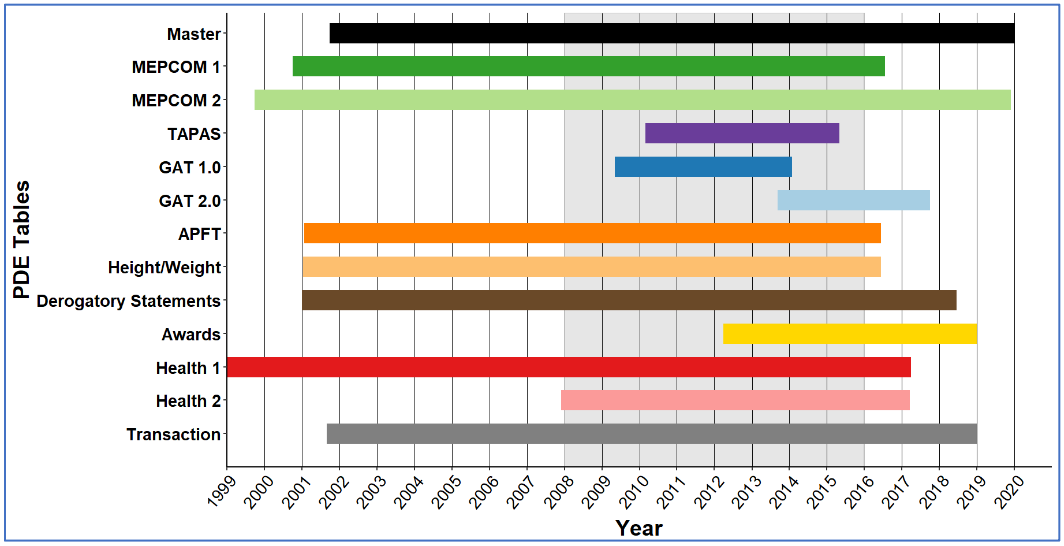
Note. First date and last date represent the time period of coverage for data tables at the time of analysis.

Sample Selection Criteria and Demographics

We selected subjects from the Master File who accessed to the U.S. Army from 2008 to 2015. This timespan was selected because it represented the best overlap of the data sources available (see Exhibit 5). Subjects were active-duty Soldiers who were initially enlisted as enlistees (n = 499,807) or commissioned as officers (n = 33,510); warrant officers (n = 446) were excluded from analyses given their low numbers and their insular role in the Army. The total sample consisted of 533,317 Soldiers with 499,807 (93.72%) being enlisted and 33,510 (6.28%) being officers (see Exhibit 6). Of the total sample, around 47.53% were in a combat arms Military Occupational Specialty (MOS), 11.63% were in a combat support MOS, and 24.39% were in a combat service support MOS. A total of 84.45% were male and 15.55% were female. The sample was 73.35% White, 20.20% Black, 4.61% Asian, 0.80% American Indian or Alaskan Native, 0.39% Native Hawaiian or Pacific Islander, and 0.65% Mixed Race/Other. The average age upon accession was 22.01 (SD = 4.24) years. These numbers are consistent with those reported annually by the DOD demographic reports (Department of Defense, 2018).

Exhibit 6.

Coverage of Data Tables Used in the Person-Event Data Environment (PDE)



Note. Shaded region represents sample selection time frame.

Exhibit 7.
Sample Characteristics with Comparisons Between Rank Groups and Missing Data

Sample Characteristics		Enlisted and Officer	Enlisted	Officer	GAT	TAPAS
Sample Size (*N*)		533,317	499,807	33,510	349,455	181,121
Rank Group (%)						
	Enlisted	93.72	100	0	92.92	99.43
	Officer	6.28	0	100	7.08	0.57
MOS Type (%)						
	Combat Arms	47.53	47.74	44.44	46.80	48.82
	Combat Support	11.63	11.69	10.73	11.94	10.48
	Combat Service Support	24.39	24.99	15.42	24.48	24.46
	Special Service	10.26	9.04	28.49	10.58	9.15
	Operations	0.12	0.12	0.11	0.13	0.02
	Unknown/NA	6.08	6.43	0.81	6.08	7.07
Accession Year (%)						
	2008	13.85	13.52	18.71	14.10	0.03
	2009	13.01	12.57	19.59	13.60	0.03
	2010	14.61	14.39	17.95	15.15	6.59
	2011	12.33	11.89	18.93	12.09	14.57
	2012	11.84	11.84	11.97	11.17	19.09
	2013	12.95	13.39	6.45	13.25	24.05
	2014	10.44	10.90	3.52	9.61	19.88
	2015	10.96	11.50	2.86	11.04	15.77
Age at Accession: *M* (*SD*)		22.01 (4.24)	21.90 (4.16)	23.59 (5.06)	22.16 (4.35)	21.51 (3.50)
Soldier Sex (%)						
	Male	84.45	84.93	77.33	84.36	84.70
	Female	15.55	15.08	22.67	15.64	15.30
Soldier Race (%)						
	White	73.35	72.96	79.07	72.87	71.98
	Black or African-American	20.20	21.08	7.06	20.66	21.66
	Asian	4.61	4.47	6.73	4.65	5.13
	American Indian/Alaskan Native	0.80	0.82	0.51	0.75	0.82
	Native Hawaiian/Pacific Islander	0.39	0.40	0.24	0.39	0.36
	Mixed Race/Other	0.65	0.26	6.40	0.67	0.05

Note. GAT = Global Assessment Tool, TAPAS = Tailored Adaptive Personality Assessment System; GAT and TAPAS columns represent characteristics of Soldiers with available data for those data sources.

Variable Selection

Variables Selected

Once conceptual and methodological profiling was complete on the full corpus of available data, variables were selected for Phase 1 modeling. The selection of variables was informed by prior research from the literature review on performance in Army and non-military contexts as well as our conceptual model of performance based on this literature. The first selection criteria were that each variable (predictor or outcome) be in a data table that included a PID so that they could be linked with other data sources. Then, using the conceptual variable categorization (described above), variables were selected that were consistent with prior research and our conceptual model of performance. Given the large pool of possible predictor variables available from various sources in the PDE, we narrowed our search to those that fit categories describing psychosocial characteristics of Soldiers. Selected predictor variables were mostly trait (stable) characteristics of Soldiers like demographic attributes (e.g., sex, race) and personality (e.g., cooperativeness, optimism). In total, 42 predictor variables were selected to test relationships with performance-related outcomes. Outcome variables were selected that best fit with the indicators of the four performance components purposed by Koopmans and colleagues (2011) as categorized the conceptual profiling process described above. In total, seven performance-related outcome variables were selected. These variables included aspects of:

1. Task performance (i.e., physical fitness test scores).
2. Counterproductive performance (i.e., alcohol abuse scores, recorded counts of bad conduct).
3. Contextual performance (i.e., recorded counts of military awards).
4. General performance (i.e., speed of promotion; see Hosek & Mattock, 2004); and,
5. Other outcomes of interest to the Army (i.e., first-term attrition, character of service upon separation).

Furthermore, some of the outcome variables represented career aggregate scores which were derived by averaging scores over time (i.e., alcohol abuse scores, physical fitness scores), while others represented an indicator using a single time point (e.g., character of service).

Variables Not Selected

The goal of the Phase 1 modeling was to gain initial insight into the simple relationships between predictor variables and outcome variables thought to be related to individual work performance. Given that there are over 4,000 variables currently available for study for our project in the PDE, we targeted variables specifically for their potential relationship with performance based on our conceptual framework for the Phase 1 modeling. These analyses were not intended to be exhaustive nor comprehensive of all the available variables contained within the PDE data sources. As such, some variables that had been studied in past research were not included in the current set of analyses. For example, we did not include variables like waivers (potential unreliability in records), courses completed (uniformity due to Army standards for promotion), weapons qualification results (dependent upon MOS, weapon type, and little variability in score categories), or unit-level constructs (e.g., unit climate).

Variables

For analyses, variables were identified to serve as predictors or performance-related outcomes (see Exhibit 12 for a summary). From data tables in the PDE, variables were either 'used' in their unaltered states or ‘derived’ through calculation and/or combination with other variables (see the “Benchmarking Crosswalks of Army data from Multiple Sources” for example R code of derived variables; Ratcliff et al., unpublished).

Predictor Variables

In total, 42 variables were identified to predict performance. Variables consisted of Army functioning variables (e.g., Rank Group), demographic variables (e.g., Soldier Sex, Soldier Race), assessment variables (e.g., cognitive ability, health functioning), measures of psychosocial traits (e.g., Adaptability, Depression), and personality traits (e.g., Adjustment, Physical Conditioning). Below each variable is described in detail. For a correlation analysis between predictor variables, see Exhibit 14.

Rank Group (RANK_PDE_GRP | Derived | Dichotomous). Rank Group represented the group in which the Soldier’s rank fell under and was coded as a categorical variable with two levels: enlisted or officer. This variable (RANK_PDE_GRP) was derived from the Master table using the Soldier’s first rank (RANK_PDE) as determined by the file date (FILE_DT). Soldiers were coded as ‘enlisted’ if their first rank included PV1 (Private), PV2 (Private), PFC (Private First Class), CPL (Corporal), SPC (Specialist), SGT (Sergeant), SSG (Staff Sergeant), SFC (Sergeant First Class), EEE (Senior Enlisted: Master Sergeant and higher). Soldiers were coded as ‘officer’ if their first rank included 2LT (Second Lieutenant), 1LT (First Lieutenant), CPT (Captain), MAJ (Major), OOO (Senior Officer: Lieutenant Colonel and higher).

MOS Type (MOS_TYPE | Derived | Categorical). MOS Type represented the categorization of the primary purpose of a Soldier’s MOS and was coded as a categorical variable with six levels: combat arms, combat support, combat service support, special service, operations, and unknown. This variable was derived from the Master, MEPCOM 1, and MEPCOM 2 tables using the variables PRI_SVC_OCC_CD, MOS, and ACC_PRI_SVC_OCC_CD, respectively. For each variable, the first three digits were taken to represent the MOS. Next, a new combined variable (MOS.CB) was created which coalesced the three together with preference to the ordering presented above. Finally, using a PDE lookup table (LKUP_ARMY_MOS) that categorized MOS types by rank group using the six levels presented above, a MOS_TYPE variable was created.

Soldier Sex (PN_SEX_CD | Used | Dichotomous). The reported sex of the Soldier was used as a categorical variable with two levels: male or female. This variable (PN_SEX_CD) was taken from the Master table.

Soldier Race (RACE_CD_RE | Derived | Categorical). The reported race of the Soldier was derived as a categorical variable with six levels: White, Black, Asian, American Indian/Alaskan Native, Native Hawaiian/Pacific Islander, or Other. This variable RACE_CD_RE was derived from the RACE_CD variable found in the Master table. Using a PDE lookup table (LKUP_DMDC_CIV_RACE_CD) the 32 racial categories found in RACE_CD were reduced such that all mixed race categories (e.g., Black or African-American/White) were collapsed into a single ‘Mixed Race/Other’ category.

Age at Accession (AGE_ACC | Calculated | Numeric). The age at accession was derived as a numeric variable representing the age (in years) of Soldiers when they joined the Army (i.e., accession). To calculate values for this variable, a variable representing the date of accession first needed to be derived.

The accession date was derived from the MEPCOM 1, Transaction, and Master tables using the variables DATE_ACC (actual date of accession for enlisted), AFMS_BASE_DT (first date of military service), and USVC_INIT_ENT_DT (initial enlistment date), respectively. A new combined variable (DATE_ACC.CB) was created which coalesced the three together with preference to the ordering presented above. It should be noted that for records after 2001, the first file date (FILE_DT) in the Master table is fairly accurate (within a couple of weeks or months of the accession date); dates prior to 2001 over inflate number of people due to the conversion to electronic records around 2001.

Next, to calculate age at accession the birth date of Soldiers was needed. The birth date was derived from the Master, MEPCOM 1, and MEPCOM 2 tables using the variables DATE_BIRTH_PDE, DATE_BIRTH, DATE_BIRTH, respectively. A new combined variable (DATE_BIRTH.CB) was created which coalesced the three together with preference to the ordering presented above.

Lastly, the age at accession variable (AGE_ACC) was created by calculating the numeric length of time (in years) from the Soldier’s date of birth to their date of accession.

AFQT Score (AFQT_PCTL.CB | Derived | Numeric). The Armed Forces Qualification Test (AFQT) Score was derived as a numeric variable representing the percentile score the Soldier received on the AFQT. The AFQT is a component of the Armed Services Vocational Aptitude Battery (ASVAB) that assesses Soldiers cognitive ability for use in placement into various military jobs. The AFQT score represents a composite percentile score of a Soldier for four broad subject-specific sections (e.g., word knowledge, paragraph comprehension, arithmetic reasoning, and mathematics knowledge). A minimum AFQT score of 31 is required to enlist in the Army. The AFQT Score was derived from the MEPCOM 2, Master, and MEPCOM 1 tables using the variables AFQT_PCTL_SCR_QY, AFQT_PCTL_SCR_QY, AFQT, respectively. A new combined variable (AFQT_PCTL.CB) was created which coalesced the three together with preference to the ordering presented above. Lastly, all scores of ‘0’ were recoded as ‘NA’ as these scores were not possible for Army accession requirements and were likely coded erroneously.

PULHES Score (PULHES.MEAN | Calculated | Numeric). The PULHES Score was derived as a numeric variable representing the initial health of Soldiers entering the Army. The Army uses PULHES as shorthand within a Physical Profile Serial System to qualify the physical capacity of a Soldier’s body systems. The requirements for PULHES vary by MOS. PULHES stands for six different health screening tests that are used to indicate overall functional capacity: Physical condition (P), Upper extremities (U), Lower extremities (L), Hearing (H), Eyes (E), and Psychiatric (S). For each screening area, a four-point scale is used to indicate the functional capacity of a particular organ or system in the body:

(a) 1 = High level of medical fitness;
(b) 2 = Some medical condition or physical defect that may require some activity limitations;
(c) 3 = One or more medical conditions or physical defects that may require significant limitations;
(d) 4 = One or more medical conditions or physical defects of high severity that limits military duty.

For example, a code of '1' or '2' for a given health category typically means that the Soldier can deploy and a '3' or '4' means that the Soldier cannot deploy (see Department of the Army, 2019b, 2019d).

The PULHES Score was derived from the MEPCOM 1 table using the variable PULHES (six digits split into individual variables: PULHES.P, PULHES.U, PULHES.L, PULHES.H, PULHES.E, PULHES.S) and the MEPCOM 2 table using the variables PHY_CPCY_PHY_PRFL_MOD_CD, UXTR_PHY_PRFL_MOD_CD, LXTR_PHY_PRFL_MOD_CD, HRG_PHY_PRFL_MOD_CD, VSN_PHY_PRFL_MOD_CD, and PSYC_PHY_PRFL_MOD_CD. Next, combined versions of each health indicator were created by coalescing the six variables from each MEPCOM table with preference given to MEPCOM 2. Finally, an overall composite variable was created (PULHES.MEAN) representing the mean across the six health screening assessments.

The Global Assessment Tool (GAT). The Global Assessment Tool (GAT; currently known as the Azimuth Check) was designed to serve as the conduit for annual self-assessment of resilience-related characteristics for Soldiers and their families. The GAT contains a constellation of psychosocial measures that map onto four major components:

- *Emotional fitness* (e.g., adaptability, character, depression) reflects one's mood, satisfaction, coping styles, and character strengths.
- *Social fitness* (e.g., loneliness, organizational trust) is indicative of one's feelings towards close friends, unit members, and leaders.
- *Family fitness* (e.g., family satisfaction and support) is an indicator of one's personal relationships with family and romantic partners.
- *Spiritual fitness* (i.e., life meaning) reflects the degree to which one ascribes meaning and purpose to one's life and world around them. It is important to note that the term

'spiritual' is not indicative of any sort of religiosity or religious belief but an existential meaning one draws from their life and the world around them.

- A fifth component, *physical fitness*, was added later in version 2.0 in 2014 which assesses aspects of nutrition, sleep habits, and substance abuse (see Lester et al., 2015; Peterson et al., 2011; Vie et al., 2016).

Thirteen GAT variables were taken from the two GAT tables 1.0 and 2.0 which represented two versions of the GAT administered to consenting Soldiers from 2009 to 2017 (GAT 1.0: 2009–2014; GAT 2.0: 2013–2017). For the current research, we did not include measures from the Family Fitness component or the Physical Fitness component. Family Fitness measures were dependent upon whether a Soldier had a family to reference, and the Physical Fitness measures were not assessed in a way conducive of forming scales (e.g., items were not interchangeable). The GAT is typically measured on an annual basis for Soldiers, however, we only used a single time occasion (i.e., the first measurement point). This decision followed from prior research which indicated that the GAT measures showed mostly trivial magnitudes of change over time, suggesting the measures to be more trait-like than measuring time-varying states (Ratcliff et al., unpublished). For each version of the GAT (1.0 or 2.0), items were recoded such that higher numbers indicated stronger prevalence of the construct. Then, for each scale within each GAT version, individual items were averaged together to form a composite (e.g., adapt.scale, adapt.scale2). Finally, for each of the thirteen measures, a new combined variable (e.g., adapt.scale.CB) was created which coalesced the two GAT versions together with preference given to GAT 1.0 as it would be the earliest for each individual. Below we describe each of the 13 measures in terms of operational definition, example items, and scaling.

Adaptability (adapt.scale.CB | Calculated | Scale). Adaptability measured the ability to alter one's course and perceived cognitive flexibility. Using the stem "How well do these statements describe you", example items include "I am good at changing myself to adjust to changes in my life" or "I can usually fit myself into any situation." Adaptability items (three items for both GAT 1 and 2) were evaluated using a 5-point Likert scale (1 = *Not like me at all*, 5 = *Very much like me*).

Active Coping (acope.scale.CB | Calculated | Scale). Active Coping measured strategies that involve planning or taking directed action. Using the stem "How well do these statements describe you", example items include "When something stresses me out, I try to solve the problem" or "I control my emotions by changing how I think about things". Active coping items (five items for both GAT 1 and 2) were evaluated using a 5-point Likert scale (1 = *Not like me at all*, 5 = *Very much like me*).

Passive Coping (pcope.scale.CB | Calculated | Scale). Passive Coping measured strategies that involve venting or displacement and disengagement. Using the stem "How well do these statements describe you", example items include "I usually keep my emotions to myself" (reverse-coded) or "When something stresses me out, I try to avoid it or not think about it" (reverse-coded).

Passive coping items (three items for both GAT 1 and 2) were evaluated using a 5-point Likert scale (1 = *Not like me at all*, 5 = *Very much like me*).

Character (chr.scale.CB | Calculated | Scale). Character measured strengths within the virtues of wisdom, courage, humanity, justice, temperance, and transcendence. Using the stem "Think about how you have acted in actual situations during the past four weeks", character items were designed to measure the character strengths of individuals across six character virtues. These virtues included knowledge (e.g., "critical thinking, open-mindedness, or good judgment"), courage (e.g., "persistence"), humanity (e.g., "kindness or generosity to others"), justice (e.g., "fairness"), temperance (e.g., "modesty or humility"), and transcendence (e.g., "gratitude"). Character items (24 for GAT 1 and 18 for GAT 2) were evaluated using an 11-point Likert scale (0 = *Never*, 10 = *Always*).

Catastrophizing (catastro.scale.CB | Calculated | Scale). Catastrophizing measured internal explanatory style of attributions towards negative events. Using the stem "Answer in terms of how you usually think", example items include "When bad things happen to me, I expect more bad things to happen" or "When I fail at something, I give up all hope". Catastrophizing items (seven items for GAT 1 and three for GAT 2) were evaluated using a 5-point Likert scale (1 = *Not like me at all*, 5 = *Very much like me*).

Depression (depress.scale.CB | Calculated | Scale). Depression measured the prevalence of depressive symptoms of feeling down, depressed, or hopeless. Using the stem "In the past four weeks, how often have you been bothered by any of the following problems", example items include "Feeling down, depressed, or hopeless" or "Little interest or pleasure in doing things". Depression items (ten items for both GAT 1 and 2) were evaluated using a 5-point Likert scale (1 = *Not at all*, 5 = *Every day*).

Optimism (optimism.scale.CB | Calculated | Scale). Optimism measured the generalized expectation for positive future events. Using the stem "Answer according to your own feelings, rather than how you think most people would answer", example items include "In uncertain times, I usually expect the best" or "I rarely count on good things happening to me" (reverse-coded). Optimism items (four items for both GAT 1 and 2) were evaluated using a 5-point Likert scale (1 = *Strongly disagree*, 5 = *Strongly agree*).

Positive Affect (posaffect.scale.CB | Calculated | Scale). Positive Affect measured subjective feelings of positive affect. Using the stem "How often have you felt this way during the past four weeks", example items include those in a general dimension of positive affect (e.g., "Inspired"), and those across more specific dimensions of positive affect like self-assurance (i.e., "Proud"), joviality (e.g., "Joyful"), and serenity (e.g., "Peaceful/calm"). Positive affect items (ten items for GAT 1 and nine for GAT 2) were evaluated using a 5-point Likert scale (1 = *Never*, 5 = *Most of the time*).

Negative Affect (negaffect.scale.CB | Calculated | Scale). Negative Affect measured Subjective feelings of negative affect. Using the stem "How often have you felt this way during the past four weeks", example items include those in a general dimension of negative affect (e.g., "Upset"), and those across more specific dimensions of negative affect like fear (i.e., "Scared/fearful"), hostility (e.g., "Hostile"), guilt (e.g., "Ashamed"), and sadness (e.g., "Sad"). Negative affect items (11 items for GAT 1 and nine for GAT 2) were evaluated using a 5-point Likert scale (1 = *Never*, 5 = *Most of the time*).

Loneliness (lone.scale.CB | Calculated | Scale). Loneliness measured feelings of being alone and separated from others. Using the stem "Please be honest as possible", example items include "How often do you feel left out" or "How often do you feel close to people" (reverse-coded). Loneliness items (three items for both GAT 1 and 2) were evaluated using a 5-point Likert scale (1 = *Never*, 5 = *Most of the time*).

Organizational Trust (orgtrust.scale.CB | Calculated | Scale). Organizational Trust measured trust in the organization (i.e., peers, leaders) in terms of ability, benevolence, and integrity. Using the stem "Please indicate how strongly you agree or disagree with each of the following statement", example items include those pertaining to facets of ability (e.g., "My immediate supervisor has much knowledge about the work that needs to be done"), benevolence (i.e., "I trust my fellow Soldiers in my unit to look out for my welfare and safety"), and integrity (e.g., "Overall, I trust my immediate supervisor"). Organizational trust items (five items for GAT 1 and four for GAT 2) were evaluated using a 5-point Likert scale (1 = *Strongly disagree*, 5 = *Strongly agree*).

Work Engagement (wkengage.scale.CB | Calculated | Scale). Work Engagement measured satisfaction and commitment to work. Using the stem "How well do these statements describe your feelings about your job", example items include "My work is one of the most important things in my life" or "I am committed to my job." Work engagement items (four items for both GAT 1 and 2) were evaluated using a 5-point Likert scale (1 = *Not like me at all*, 5 = *Very much like me*).

Life Meaning (lifemean.scale.CB | Calculated | Scale). Life Meaning measured sense of purpose and meaning to life and work. Using the stem "Answer in terms of whether the statement describes how you actually live your life", example items include some worded generally "My life has meaning" or "I believe there is a purpose for my life" and others worded for a military-specific context (e.g., "The job I am doing in the military has enduring meaning"). Life meaning items (three items for both GAT 1 and 2) were evaluated using a 5-point Likert scale (1 = *Not like me at all*, 5 = *Very much like me*).

Tailored Adaptive Personality Assessment System (TAPAS). The Tailored Adaptive Personality Assessment System (TAPAS) was designed as an assessment of 22 personality traits to be used in initial assessments of Soldiers for the purposes of appropriate job placement. The personality traits in the TAPAS broadly map onto the Big Five inventory with one, Physical Conditioning, added for its relevance to the Army domain. Using an item response theory approach

(IRT), the TAPAS is administered to Soldiers upon accession as a computer adaptive assessment. During the assessment, Soldiers are forced to choose between pairs of personality statements that are matched on levels of social desirability and extremity (Stark et al., 2014). For example, in one trial, Soldiers are presented with two statements "I get along well with coworkers" (cooperation) and "I am known as a 'quick thinker'" (intellectual efficiency) and are asked to "choose the statement that is more like you." This Multidimensional Pairwise Preference (MUPP) approach has several advantages like reducing the instance of faking and offering a large pool of items for the assessment to draw from. Once a Soldier has completed the assessment, IRT-based trait scores (θ, thetas) are calculated for each personality dimension. The theta parameter represents the magnitude of the latent trait or attribute being tested for the individual and is usually normed or standardized such that '0' represents the mean of the latent trait and, like a *z*-score, '1' represents the standard deviation of the calibrated sample. Moreover, theta is not analogous to a raw score commonly used in classical test theory, rather, theta represents the highest point for the estimate of maximum likelihood estimates of theta for every item. Below we describe the operational definitions of each of the 22 TAPAS measures taken from the TAPAS table which represents consented data only (for a review, see Stark et al., 2014).[2]

Achievement (ACHVMNT_THETA_SCR_QY | Used | Numeric). Achievement measures an individual's level of ambition, confidence, resourcefulness and industry.

Adjustment (ADJ_THETA_SCR_QY | Used | Numeric). Adjustment measures an individual's reaction to new situations including levels of nervousness, apprehension, anxiety and certainty.

Adaptation (ADPT_CMPS_SCR_QY | Used | Numeric). Adaptation measures an individual's ability to adapt to new problems and situations.

Adventure (ADV_THETA_SCR_QY | Use | Numeric). Adventure measures an individual's enjoyment of participating in extreme sports and outdoor activities.

Attention Seeking (ATTN_SEEK_THETA_SCR_QY | Used | Numeric). Attention Seeking measures an individual's tendency towards shyness or need for social attention, capturing boastfulness or diffidence.

Commitment to Serve (CMTS_THETA_SCR_QY | Used | Numeric). Commitment to Serve measures an individual's level of identification with the military and their strength of desire to serve their country.

[2] It should be noted that the TAPAS data might include respondents who never joined the military, and thus, might not be fully represented in analyses.

Cooperation (COOPR_THETA_SCR_QY | Used | Numeric). Cooperation measures an individual's agreeableness, trust, skepticism and suspicion or the extent they are easy or difficult to get along with.

Courage (COUR_THETA_SCR_QY | Used | Numeric). Courage measures an individual's tendency to stand up to challenges and to not be afraid to face dangerous situations.

Dominance (DOMNC_THETA_SCR_QY | Used | Numeric). Dominance measures an individual's assertiveness or submissiveness and propensity to "take charge."

Even Tempered (EVTMP_THETA_SCR_QY | Used | Numeric). Even Tempered measures an individual's disposition to anger, hostility, calmness and stability.

Intellectual Efficiency (INTLL__EFC_THETA_SCR_QY | Used | Numeric). Intellectual Efficiency measures an individual's ability to analyze and process information, astuteness or obtuseness.

Non-Delinquency (NON_DLNQY_THETA_SCR_QY | Used | Numeric). Non-Delinquency measures an individual's tendency to be lawful and to comply with authority including propensity to follow rules and regulations.

Optimism (OPTMSM_THETA_SCR_QY | Used | Numeric). Optimism measures an individual's cheerfulness and emotional outlook, capturing positivism, negativism, depression and contentment.

Order (ORD_THETA_SCR_QY | Used | Numeric). Order measures an individual's tendency to organize tasks and activities and desire to maintain neat and clean surroundings.

Physical Condition (PHY_COND_THETA_SCR_QY | Used | Numeric). Physical Condition measures an individual's proclivity for participating in sports, physical activity and outdoor activities as well as sedentary tendencies.

Responsibility (RSBY_THETA_SCR_QY | Used | Numeric). Responsibility measures an individual's dependability, reliability, and tendency to make every effort to keep their promises.

Sociability (SCBLTY_THETA_SCR_QY | Used | Numeric). Sociability measures an individual's level of interest in social interaction including gregariousness, talkativeness and introversion.

Self-Control (SELF_CTRL_THETA_SCR_QY | Used | Numeric). Self-Control measures an individual's patience, deliberateness, caution, impulsiveness and rashness.

Situational Awareness (SITNL_AWRNS_THETA_SCR_QY | Used | Numeric). Situational Awareness measures an individual's tendency to pay attention to their surroundings and to rarely get lost or surprised.

Selflessness (SLFNS_THETA_SCR_QY | Used | Numeric). Selflessness measures an individual's selflessness and selfishness or tendency to be giving, charitable, egotistical or greedy.

Team Orientation (TEAM_ORNTN_THETA_SCR_QY | Used | Numeric). Team Orientation measures an individual's tendency to prefer working in teams and make people work together better.

Tolerance (TOL_THETA_SCR_QY | Used | Numeric). Tolerance measures an individual's acceptance of differing customs, viewpoints, persons, and events or bias and lenience towards persons and situations.

Outcome Variables

In total, seven variables were identified as related to performance or performance outcomes. Variables consisted of career-spanning performance variables (i.e., AUDIT-C Score, APFT Score, Speed of Promotion), extraordinary instances of good conduct (i.e., Awards Count) and bad conduct (i.e., Bad Papers Count), and outcomes of service (i.e., First-Term Attrition, Character of Service). Below each variable is described in detail. For a correlation analysis between outcome variables, see Exhibit 15.

AUDIT-C Score (AUDITC_TOTALSCORE_MEAN | Calculated | Numeric). The Alcohol Use Disorders Identification Test-Concise (or AUDIT-C) score was derived as a numeric variable to represent alcohol misuse. As an indicator of counterproductive work performance (cf. Koopmans et al., 2011), the AUDIT-C was designed as a screening questionnaire for alcohol misuse and abuse. Originally a 10-item measure, the most commonly used version is the 3-item measure with rating scales that go from zero to four points. The AUDIT-C items include:

(a) "How often do you have a drink containing alcohol?" (0 = *Never*; 1 = *Monthly or less*; 2 = *2–4 times per month*; 3 = *2–3 times per week*; 4 = *4+ times per week*);
(b) "How many standard drinks containing alcohol do you have on a typical day (on days that you drink)?" (0 = *1 or 2*; 1 = *3 or 4*; 2 = *5 or 6*; 3 = *7 to 9*; 4 = *10 or more*); and
(c) "How often do you have six or more drinks on one occasion" (0 = *Never*; 1 = *Less than monthly*; 2 = *Monthly*; 3 = *Weekly*; 4 = *Daily or almost daily*).

An AUDIT-C Score is calculated by summing the three items together to get a total score that ranges from 0–12 points. A total score of ≥4 points for men and ≥3 points for women indicate risky to abusive drinking behavior. Past research has found the AUDIT-C to be valid across cultural groups (e.g., race, gender) for predicting alcohol disorders (see Bush et al., 1998; Frank et al., 2008).

The AUDIT-C Score was derived from the Health 2 table using the variable PH_AUDITCSCORE which represented the total AUDIT-C score the Soldier received at a given periodic health assessment. A new variable AUDITC_TOTALSCORE_MEAN was created by calculating the mean of all AUDIT-C scores reported for a Soldier across their career.

APFT Score (APFT_TOTALSCORE_MEAN | Calculated | Numeric). The Army Physical Fitness Test (or APFT) Score was derived as a numeric variable to represent physical fitness. Given the mandatory requirement of all Soldiers in the Army to maintain standard levels of physical fitness, this variable served as an indicator of task work performance (cf. Koopmans et al., 2011). The APFT is a performance test to measure upper and lower body muscular endurance to indicate a Soldier's ability to perform physically on the job. The assessment consists of three events: push-ups, sit-ups, and a 2-mile run performed consecutively on the same day with a maximum of 20 minutes' rest between events (Army ROTC, n.d.). For each event, Soldiers must obtain a score of at least 60 points and an overall score of at least 180 points out of 300; APFT standards are adjusted for age and gender (Department of the Army, 2012). An overall APFT score represents an averaged composite score of the three events.

The APFT Score was derived from the APFT table using the variable SCORE_TOTAL which represented the total APFT score the Soldier received at a given assessment. A new variable APFT_TOTALSCORE_MEAN was created by calculating the mean of all APFT scores reported for a Soldier across their career.

Award Count (award_count | Derived | Integer). The Award Count variable was derived as an integer variable to represent the awards a Soldier received across their career. Given that awards are typically given to Soldiers who go above and beyond their daily duties (Department of the Army, 2019a), the Award Count variable was categorized as an indicator of contextual work performance (cf. Koopmans et al., 2011). The Award Count variable was derived from the Awards table using the variable AWRD_CD which represented a code for an award received at a given date (AWRD_DT). A new variable award_count was created by summing a raw count of awards each Soldier had received.

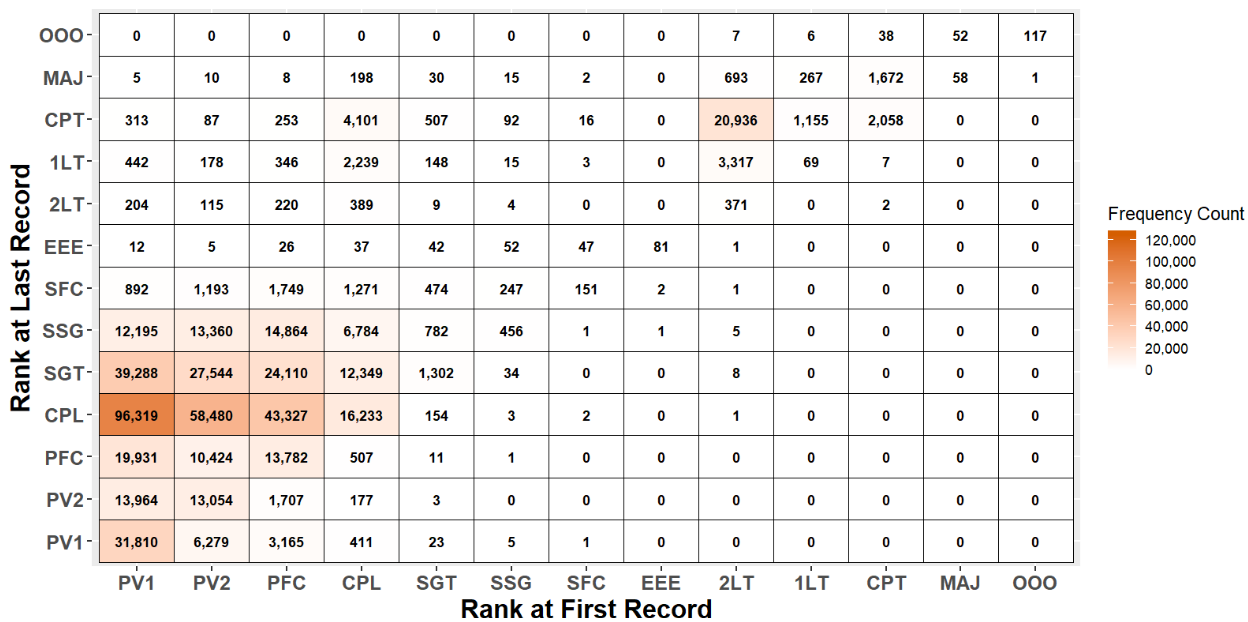
Bad Paper Count (badpaper.overall | Calculated | Integer). The Bad Paper Count variable was derived as an integer variable to represent the number of bad papers or documents of bad conduct that a Soldier received across their career. Given that bad papers are given to Soldiers for conduct that runs against Army goals (Department of the Army, 2016), the Bad Paper Count variable was categorized as an indicator of counterproductive work performance (cf. Koopmans et al., 2011). The Bad Paper Count variable was derived from the Derogatory Statements table using the variable NAME_DEROG_DOC which represented a code for a derogatory statement received at a given date (DATE_DEROG_EFF). Three new variables—COURT_MARTIAL, LETTER_REPRIMAND, and ARTICLE15— were created to represent a total count of specific derogatory statements a Soldier had received across their career (as determined by the codes:

“COURT MART”, “LTR REPR”, and “DA 2627” on the NAME_DEROG_DOC variable). Given extremely low base rates for article 15s (1.31% > 0), letters of reprimand (1.98% > 0), and courts martials (0.47% > 0), these variables were combined into an overall aggregate. A new variable, badpaper.overall, was created by summing across the total number of court martials (COURT_MARTIAL), letters of reprimand (LETTER_REPRIMAND), and article 15s (ARTICLE15) each Soldier had received.

Speed of Promotion (SOP.RANK_HIGH.STDZ2 | Calculated | Numeric). The Speed of Promotion variable was derived as a numeric variable to represent the time taken to achieve a Soldier’s highest rank. Given that speed of promotion is a general proxy for a Soldier’s performance, the Speed of Promotion variable was categorized as an indicator of general work performance (cf. Koopmans et al., 2011). The Speed of Promotion variable was derived from the Master table along with the date of accession variable derived earlier (DATE_ACC.CB). The data were then ordered by the file date from the Master table (FILE_DT) to pull out a Soldier’s first and last recorded rank as new variables (RANK_FIRST, RANK_LAST). The two rank variables were then ordered as a factor from PV1 (Private) to OOO (Senior Officer) in line with the progressive rank structure of the Army (see Exhibit 7).

Exhibit 8.

Frequency Count of Soldier’s First Recorded Rank and Last Recorded Rank



Note. PV1 = Private, PV2 = Private, PFC = Private First Class, CPL = Corporal/Specialist, SGT = Sergeant, SSG = Staff Sergeant, SFC = Sergeant First Class, EEE = Senior Enlisted (Master Sergeant and higher), 2LT = Second Lieutenant, 1LT = First Lieutenant, CPT = Captain, MAJ = Major, OOO = Senior Officer (Lieutenant Colonel and higher).

Next, a variable RANK_HIGH was created to represent the highest last rank for a Soldier along with the date of said rank (RANK_LAST_DT). A set of demotion variables were then created to indicate if the last rank was less than or equal to the first (RANK_DEMOTE_LOW.EQ) and if the highest rank was equal to the first rank (RANK_DEMOTE_EQ). This was done so that those who never got promoted (same first and last rank) would be excluded (given NAs) in the speed of promotion calculation since these individuals never got promoted. The next step involved providing an exact calculation (in months) of the time from a Soldier's accession date until their date of promotion to their highest rank (SOP.RANK_HIGH).[2] For a cross-tabulation of time to promotion by starting rank and highest rank of record, see Exhibit 8.

Exhibit 9.

Average Time to Promotion (in Months) by Rank at First Record and Highest Rank of Record (SDs in Parentheses)

Highest Rank of Record \ Rank at First Record	PV1	PV2	PFC	CPL	SGT	SSG	SFC	EEE	2LT	1LT	CPT	MAJ	OOO
OOO	NC	NC	NC	NC	NC	NC	NC	NC	99.14 (25.20)	93.00 (24.53)	70.20 (46.60)	50.98 (31.01)	—
MAJ	123.60 (27.25)	133.90 (4.95)	127.88 (22.54)	132.43 (17.60)	131.33 (16.53)	133.40 (6.68)	134.50 (6.36)	NC	124.85 (16.35)	102.08 (20.08)	79.50 (26.72)	—	—
CPT	70.06 (22.27)	87.32 (32.60)	91.45 (28.95)	60.01 (15.67)	54.67 (11.50)	54.82 (8.71)	66.19 (17.31)	NC	69.78 (23.06)	18.52 (19.96)	—	—	—
1LT	75.75 (22.37)	98.13 (26.20)	92.47 (24.83)	39.96 (23.97)	28.85 (16.61)	38.67 (30.25)	34.00 (8.54)	NC	24.76 (12.48)	—	—	—	—
2LT	68.38 (29.18)	98.90 (23.85)	87.85 (25.29)	54.57 (30.53)	63.78 (47.66)	39.75 (39.34)	NC	NC	—	—	—	—	—
EEE	97.42 (25.03)	95.00 (25.65)	99.58 (29.68)	106.03 (30.81)	93.12 (42.03)	81.00 (39.72)	55.70 (30.53)	—	—	—	—	—	—
SFC	116.89 (17.99)	116.29 (16.89)	111.46 (17.32)	105.51 (19.10)	92.69 (27.92)	60.64 (30.58)	—	—	—	—	—	—	—
SSG	84.99 (19.59)	85.82 (18.67)	82.82 (19.17)	70.70 (23.48)	50.78 (27.58)	—	—	—	—	—	—	—	—
SGT	53.14 (14.19)	52.79 (15.17)	51.84 (15.91)	40.43 (15.12)	—	—	—	—	—	—	—	—	—
CPL	25.54 (4.35)	25.16 (4.32)	24.29 (4.90)	—	—	—	—	—	—	—	—	—	—
PFC	14.91 (4.91)	13.83 (5.13)	—	—	—	—	—	—	—	—	—	—	—
PV2	8.35 (3.12)	—	—	—	—	—	—	—	—	—	—	—	—
PV1	—	—	—	—	—	—	—	—	—	—	—	—	—

Rank at First Record

Note. PV1 = Private, PV2 = Private, PFC = Private First Class, CPL = Corporal/Specialist, SGT = Sergeant, SSG = Staff Sergeant, SFC = Sergeant First Class, EEE = Senior Enlisted (Master Sergeant and higher), 2LT = Second Lieutenant, 1LT = First Lieutenant, CPT = Captain, MAJ = Major, OOO = Senior Officer (Lieutenant Colonel and higher); NC = No cases.

Lastly, a Soldier's time taken to promotion to their highest rank was standardized in relation to others who reached the same rank. This standardization controls for the unique and, at times, congressionally-mandated rules on how quickly Soldiers can move through the ranks. Therefore, the SOP.RANK_HIGH variable, representing the time (in months) taken to achieve a given Soldier's highest rank, was used to calculate means and standard deviations for all pairings of starting and highest ranks on record. Values obtained by this cross-tabulation were used to standardize the final Speed of Promotion variable (SOP.RANK_HIGH.STDZ2).

First-Term Attrition (ATTRIT_FIRST_TERM | Calculated | Categorical). The First-Term Attrition variable was derived as a dichotomous categorical variable to represent whether a Soldier completed their first-term contractual obligations with the Army. Although attrition did not cleanly map onto any of the performance dimensions proposed by Koopmans and colleagues (2011), we included this variable because it is an important outcome of interest to the Army (Department of the Army, 2019c, 2020). The First-Term Attrition variable was derived from the two MEPCOM tables (MEPCOM 1, MEPCOM 2) and the Transaction table.

For both MEPCOM tables, there was a variable that indicated the length of term (in months) of a Soldier's first term in the Army upon accession (TERM and ACC_ENLM_AGMT_DRTN_YR_QY, respectively). These two variables were combined to create a term length variable (TERM.CB) with preference given to the term variable in the MEPCOM 1 table. Next, a variable was created (TERM_END_DT) to indicate the date of the end of a Soldier's first term based on the addition of the length of a Soldier's first term to the accession date (DATE_ACC.CB).

To calculate the date at which a Soldier separated from the Army, the Transaction table was used along with a transaction code variable (ADP_TXN_TYP_CD) and a variable representing the date of the last recorded transaction (TXN_EFF_DT). First, an attrition variable was created (ATTRIT.CB) by recoding the transaction codes to be indicative of an attrition loss event (e.g., "Loss to Civilian Life", "Loss to Retired") or a non-attrition admin/gain event (e.g., "Name Change", "Immediate Reenlistment"). For a full mapping of the recoding, see our "Benchmarking Crosswalks of Army data from Multiple Sources" report (Ratcliff et al., unpublished[3]). Second, an attrition date variable was created (ATTRIT_DT.CB) by using the date of last transaction (TXN_EFF_DT) if the last transaction code corresponded to an attrition code. If the Soldier had not separated from the Army as of the end of our data coverage for the Transaction table (i.e., 2018-12-31), a date in the future (i.e., "2020-01-01") was assigned for those Soldiers.

The final step to create a First-Term Attrition variable (ATTRIT_FIRST_TERM) involved creating a dichotomous categorical variable that indicated whether a Soldier had attrited before their first term was up or had not. This calculation was performed by assigning a '1' to Soldiers whose attrition date (ATTRIT_DT.CB) was less than or equal to their date of attrition (ATTRIT_DT.CB) and a '0' if otherwise.

Character of Service (CHAR_SVC_CD2 | Derived | Categorical). The Character of Service variable was derived as a dichotomous categorical variable to represent whether a Soldier was discharged with an 'honorable' designation when separating from the Army or not. Although character of service did not cleanly map onto any of the performance dimensions proposed by Koopmans and colleagues (2011), we included this variable because it is an important outcome of interest to the Army (Department of the Army, 2019c, 2020). To derive this variable, the character

[3] This publication has been approved for distribution by ARI pending revisions from the SDAD research team.

of service code (CHAR_SVC_CD) a Soldier receives upon discharge was used from the Transaction table. There are ten possible codes that can be given using the character of service variable. We recoded this variable into a new one (CHAR_SVC_CD2) to represent codes that either represent 'honorable' service (e.g., "honorable", "under honorable conditions (general)", "honorable for VA purposes") or 'not honorable' service (e.g., "bad conduct", "under other than honorable", "dishonorable"). For a full mapping of the recoding, see our "Benchmarking Crosswalks of Army data from Multiple Sources" report (Ratcliff et al., unpublished).

Benchmarking our work

As an initial check of our efforts, we benchmarked our sample against values reported in several publicly available reports. We found our top-level sample characteristic numbers and the observed trends over time were consistent with those reported annually by the DOD demographic reports (DOD, 2003, 2010, 2018, 2019) for active-duty Army strength. Our counts of annual accessions were also in keeping with the values reported from 2000 to 2016 by other DOD-sponsored researchers operating in the PDE (Knapp et al., 2018b). This benchmarking effort used a variety of sources, including the Active Duty Personnel Master File, Military Entrance Processing Command (MEPCOM) tables, and Active Duty Military Personnel Transaction File covering a time range from September 1999 to December 2019. Data were examined by cohort year (indicating a Solider was in the Army for a given year) and by accession year (indicating the year a Soldier accessed to the Army). These findings confirm both the quality of Army data available within the PDE and provide assurances that our later modeling results are grounded in an accurate depiction of the force. More information about this work was reported in our report "Benchmarking Crosswalks of Army data from Multiple Sources."

PART 3: ANALYSIS

Analysis 1: Examining the U.S. Army's Global Assessment Tool (GAT)

Prior to beginning our modeling proper, we first examined one of our chief sources of data pertaining to the social characteristics of Soldiers, the Global Assessment Tool (GAT). The GAT (now known as the Azimuth Check) is a multidimensional instrument drawn from several different original measures designed to assess characteristics related to resilience. First implemented in 2009, all active-duty Soldiers are required to complete the GAT on an annual basis. Although widely tested and found to be psychometrically sound, there has been little work done to examine its longitudinal properties. We had quickly identified the GAT as a source of potential predictors of Soldier performance, which made it more enticing because of the regular intervals at which all Soldiers were providing data across time. To understand how best to incorporate this data into our models, we conducted a thorough analysis of the Soldier responses on the GAT across time to establish how sensitive it was to changes in Soldiers' internal states.

We accessed GAT data from 2009 to 2017 covering the implementation of two different instrument versions (i.e., GAT version 1.0 and 2.0). Our sample for the two studies involved 95,277 respondents in Study 1 (GAT 1.0) and 57,771 respondents in Study 2 (GAT 2.0). Across both studies, respondents were active-duty Soldiers who were initially enlisted in the U.S. Army and provided consent to have their data used for the study. Respondents were selected with an accession date between the start of 2009 and the end of 2014 in Study 1. For Study 2, respondents were selected with an accession date between the start of 2012 and the end of 2017, which marked the end of the time period for which GAT 2.0 data was available to us at the time of our analyses. Because our research question changed over time, each respondent in our sample had to have at least two occasions with completed data. To control the total number of responses, we included only the first five occasions when a Soldier completed the GAT.

Once data had been linked, we conducted a number of tests examining the longitudinal stability of the GAT, including: (a) repeated measures analysis of variance (RM-ANOVA); (b) repeated measures structural equation models (RM-SEM); (c) tests of measurement invariance (MI); and (d) repeated measures confirmatory factor analysis (RM-CFA). For most measures, we found that Soldiers' responses on the GAT were stable over time. However, responses to the Organizational Trust, Work Engagement, and Life Meaning scales show some meaningful variation across time.

More information about our review of the GAT's longitudinal properties has been submitted and reviewed by ARI and is available in several published reports (Ratcliff et al., 2021a; Ratcliff et al., 2021b).

Analysis 2: Relationships Analysis of Soldier Performance

Analysis Goals

Having profiled all the data available to us in the PDE for use in our analysis, we needed to decide how to proceed. We opted to take a phased approach, starting with an initial analysis that would provide us with insight into the relationship between variables while also pointing to the feasibility of developing models for testing. For our first analysis, we choose to focus on simple relationships between predictors and outcomes of performance. In effect, it acted as a proof-of-concept analysis to determine if the data available in the PDE could reasonably be used to capture the relationship between performance antecedents, indicators, and outcomes as conceptualized in our theoretical framework.

Our strategy for this phase was therefore twofold: (1) identify and test key variables of interest that could be explored in greater detail with later analyses and (2) to the degree that we detected significant relationships across variables, test the relative dominance of predictor variables on outcomes of interest. The first set of analyses consisted of a series of simple tests examining one predictor and one outcome of interest (e.g., zero-order correlation, simple linear regression, t-test, one-way ANOVA). Because the focus was on identifying the presence of basic relationships, none of the modeling performed in this phase of the project accounted for changes across time (i.e., there were no tests involving repeated measures or time-varying variables). Additionally, because of the unreliable nature of the unit-level data that we had identified during the data exploration stage, we choose to focus our work on individual performance. Following our initial analysis, we expanded on a portion of this work by using a simple logistic analysis to examine the relationship between key predictor variables and one of the performance outcomes (i.e., rate of promotion), which we expanded from a binary outcome to include three categories. This provided another proof-of-concept demonstrating administrative data can be employed to map the relationship between (certain types of) performance a more nuanced view of outcomes than is traditionally used in the Army.

Analyses

Several statistical approaches were used for data analysis at this stage of the project. All analyses were conducted in the PDE using R software (version 3.6.1).

Descriptive Statistics

A selection of descriptive statistics was first derived to provide a foundational understanding of the characteristics associated with each predictor and outcome variable. Although not comprehensive in scope, descriptive analyses of variables provide a general benchmark for the structure and range of values found for each variable as well as testing key assumptions tied to the simple relationship analysis (e.g., normality, linearity). The selection of descriptive statistics included measures of completeness (e.g., percent missing), central tendency (e.g., mean, median), dispersion (e.g., standard deviation, 95% confidence interval of mean), range (e.g., minimum and maximum values), normality (e.g., skewness, kurtosis, D'Agostino's K^2 test of normality; D'Agostino & Stephens, 1986), and reliability for composite scales (e.g., scale inter-item reliability). In addition, graphical representations of the data (e.g., boxplots, histogram/density plots, Q-Q plots, scatterplots) were examined for all the variables to examine frequencies, distributions, and simple relationships. Together, the selection of descriptive statistics and plots helped provide checks of key assumptions for each variable (for a detailed example of this process, please see the "Benchmarking Crosswalks of Army data from Multiple Sources" report (Ratcliff et al., unpublished).

Simple Relationship Analysis

For the analysis of *simple relationships,* analyses were conducted dependent on the type of predictor and outcome variables being modelled to determine the direction and the magnitude of relationships between predictor variables and outcomes.

Continuous Predictor/Continuous Outcome. Modeling of continuous predictors and continuous outcomes was performed using Pearson correlations and simple linear regressions.

Categorical Predictor/Continuous Outcome. Modeling of categorical predictors and continuous outcomes was performed using *t*-tests for predictors with two categorical levels and analysis of variance (ANOVA) for predictors with more than two categorical levels or groups. Specifically, the 'ggbetweenstats' function from the ggstatsplot package (version 0.0.12) in R was used to examine these models which was configured to test additional non-parametric models (e.g., Mann–Whitney *U* test, Kruskal–Wallis one-way ANOVA) and models robust to violations to variance heterogeneity (e.g., Welch's *t*, Heteroscedastic one-way ANOVA for trimmed means).

Continuous Predictor/Categorical Outcome. Modeling of continuous predictor variables with categorical (i.e., dichotomous) outcomes was performed using logistic generalized linear models (logit link function).

Categorical Predictor/Categorical Outcome. Modeling of categorical predictor variables with categorical (i.e., dichotomous) outcomes was performed using logistic generalized linear models (logit link function). Lastly, given the count nature of the outcome variables Award Count and Bad Paper Count, these variables were also modeled using Poisson generalized linear models (log link function) in addition to the models listed above.

Dominance Analysis

Dominance Analysis (DA; Azen & Budescu, 2003, 2006; Budescu, 1993) was conducted on the Top 7 predictor variables that had the strongest relationships with each given outcome variable as determined by the simple relationship analysis.[3] The goal of dominance analysis is to identify which predictor is more dominant (relatively important) in relation to other predictors on a given outcome (e.g., in a multiple linear regression). In dominance analysis, predictors are compared in a pairwise fashion across all possible subsets of models (e.g., zero-order models, models with two predictors, models with three predictors, up to models with all predictors being analyzed) to determine dominance based on the squared multiple correlation coefficient for each of the possible subset models (see Azen & Budescu, 2006). A predictor is considered more dominant when its relative contribution to a model (e.g., R^2) exceeds that of other predictors. For example, a predictor X_i is said to dominate X_j if the additional contribution of X_i is no less than the additional contribution of X_j (Azen & Budescu, 2006). Dominance analysis breaks dominance down into a hierarchy of dominance in the form of general, conditional, and complete dominance. General dominance represents the weighted average contribution of a predictor (e.g., R^2) across all subset models that include the predictor and those that do not. Conditional dominance represents the weighted average contribution of a predictor of all subset models of a given size (e.g., all models with two predictors). Complete dominance, the strictest level of dominance, represents cases in which a predictor's dominance holds across all possible subset models (that do not include the two predictors under comparison). Note that given the nested nature of the types of dominance, complete dominance has the property that pairwise comparisons of complete dominance will hold for the less strict levels of conditional and general dominance (see Azen & Budescu, 2006; Tang, 2014). Multi-predictor models were modelled using multiple linear regression (MLR) or multiple generalized linear models (MGLM) depending on the distribution of the outcome variable (e.g., continuous, binary). All dominance analyses were conducted with the R package 'dominanceanalysis' (version 1.0.0).

For all simple relationships analyses, standardized measures of effect size were obtained to gauge the strength of relationships (see Exhibit 9). Given the large sample sizes used in the models, often greater than 100,000, tests of significance were almost always less than an alpha of .05. Benchmarks for magnitudes of effect size are subjective; however, in the absence of prior research or guiding practical knowledge, benchmarks serve as a good baseline (Cohen, 1988).

Exhibit 10.
Measures of Effect Size and their Associated Magnitude Benchmarks

Measure of Effect Size	Typical Use Case	Trivial	Small	Medium	Large	Source
Pearson's r	Correlation	< 0.10	0.10	0.30	0.50	Cohen (1988)
Coefficient of Determination: r^2	Simple Linear Regression	< 0.01	0.01	0.09	0.25	Cohen (1988)
Coefficient of Determination: R^2	Multiple Linear Regression	< 0.02	0.02	0.13	0.26	Cohen (1988)
Cohen's d	t-test of Mean Differences	< 0.20	0.20	0.50	0.80	Cohen (1988)
Eta-squared: h^2	ANOVA	< 0.01	0.01	0.09	0.25	Cohen (1988)
Partial Eta-squared: h^2_{p}	Factorial ANOVA	< 0.01	0.01	0.06	0.14	Cohen (1988)
Generalized Eta-squared: h^2_{G}	Repeated ANOVA	< 0.02	0.02	0.13	0.26	Cohen (1988); Bakeman (2005)
Cohen's w	Proportions and Chi-square	< 0.10	0.10	0.30	0.50	Cohen (1988)
Explanatory measure of effect size: ξ	Robust t-Test & ANOVA	< 0.10	0.10	0.30	0.50	Wilcox and Tian (2011)
Odds Ratio (> 1): *OR*	Proportions and Maximum Likelihood	< 1.52	1.52	2.74	4.72	Chen et al. (2010)
Odds Ratio (< 1): *OR*	Proportions and Maximum Likelihood	> 0.66	0.66	0.37	0.21	Chen et al. (2010)

Note. ANOVA = Analysis of Variance.

Results and Discussion

Descriptive Statistics and Assumption Checks

An analysis of select descriptive statistics was conducted before modeling to gain insight into the descriptive characteristics of the predictors and outcomes related to performance. The descriptive analysis examined the following characteristics: sample counts (i.e., complete cases, missing cases), types of values (i.e., minimum/maximum value, count of unique values), measures of central tendency (i.e., mean, median), measures of dispersion (i.e., 95% confidence intervals, standard deviation), measures of normality (i.e., skewness, kurtosis, D'Agostino's K^2 test), and measures of scale reliability (i.e., omega total) for composite scale-level variables. Though not presented here (though see the "Benchmarking Crosswalks of Army data from Multiple Sources" report for examples) (Ratcliff et al., unpublished), graphical representations of variable characteristics were also reviewed for each variable (e.g., histogram/density plot, Q-Q- plot, scatterplot) for examination of assumptions. Zero-order correlations were also conducted for predictor variables and outcome variables to show strength between variables. Overall, most variables could reasonably be treated as normally distributed (for exception, see Bad Paper Count below). Due to the large sample sizes, inferential tests of normality like D'Agostino's K^2 test of normality (D'Agostino & Stephens, 1986) were almost always significant (except for Adaptation). This finding is not surprising given that inferential tests of normality become very sensitive to even small departures of normality with larger samples. Thus, these tests were viewed in light of the holistic picture formed by indicators of skewness, kurtosis, and graphical representations of variable distributions which did not indicate severe violations as the Central Limit Theorem would predict for large samples.

Predictors

The descriptive statistics for predictor variables are summarized in Exhibit 10. Overall, values (e.g., min/max, means, *SD*s) were consistent with the normal range of the predictor variables. Measures of skewness did not exceed an absolute value of three for most variables with the exception of PULHES Score (skewness = 3.79). Measures of (excess) kurtosis did not exceed an absolute value of seven for most variables with the exception of PULHES Score (Kurtosis = 33.27). Thus, most measures of skewness and kurtosis for predictor variables were within acceptable limits (±3 for skewness, ±7 for excess kurtosis; see Hair Jr et al., 2010; Kline, 2011). Scale reliabilities all exceeded the benchmark threshold of 0.70 for the GAT measures. Lastly, a correlation matrix of the predictor variables is presented in Exhibit 14. The notable pattern observed within the zero-order correlations (Pearson and point-biserial correlations) between predictors is that the GAT measures showed higher correlations amongst themselves than the other predictors. Also, the highest observed zero-order correlation (r = 0.71) was between Adaptation and Physical Condition.

Exhibit 11.

Summary of Selected Predictor and Outcome Variables

Variables	Variable Name	Type	Performance Component	Usage	Expected Range	# Items
Predictor Variables						
Accession						
Rank Group	RANK_PDE_GRP	Army Function	—	Derived	2	—
MOS Type	MOS_TYPE	Army Function	—	Derived	6	—
Soldier Sex	PN_SEX_CD	Demographic	—	Derived	2	—
Soldier Race	RACE_CD_RE	Demographic	—	Derived	6	—
Age at Accession	AGE_ACC	Demographic	—	Calculated	17– ℝ	—
AFQT Score	AFQT_PCTL.CB	Cognitive Ability	—	Derived	1–99	4
PULHES Score	PULHES.MEAN	Health	—	Calculated	1–4	6
GAT						
Adaptability	adapt.scale.CB	Psychosocial	—	Calculated	1–5	3
Active Coping	acope.scale.CB	Psychosocial	—	Calculated	1–5	5
Passive Coping	pcope.scale.CB	Psychosocial	—	Calculated	1–5	3
Character	chr.scale.CB	Psychosocial	—	Calculated	0–10	24, 18
Catastrophizing	catastro.scale.CB	Psychosocial	—	Calculated	1–5	7, 3
Depression	depress.scale.CB	Psychosocial	—	Calculated	1–5	10
Optimism	optimism.scale.CB	Psychosocial	—	Calculated	1–5	4
Positive Affect	posaffect.scale.CB	Psychosocial	—	Calculated	1–5	10, 9
Negative Affect	negaffect.scale.CB	Psychosocial	—	Calculated	1–5	11, 9
Loneliness	lone.scale.CB	Psychosocial	—	Calculated	1–5	3
Organizational Trust	orgtrust.scale.CB	Psychosocial	—	Calculated	1–5	5, 4
Work Engagement	wkengage.scale.CB	Psychosocial	—	Calculated	1–5	4
Life Meaning	lifemean.scale.CB	Psychosocial	—	Calculated	1–5	5
TAPAS						
Achievement	ACHVMNT_THETA_SCR_QY	Personality	—	Used	ℝ	varies
Adjustment	ADJ_THETA_SCR_QY	Personality	—	Used	ℝ	varies
Adaptation	ADPT_CMPS_SCR_QY	Personality	—	Used	ℝ	varies
Adventure	ADV_THETA_SCR_QY	Personality	—	Used	ℝ	varies
Attention Seeking	ATTN_SEEK_THETA_SCR_QY	Personality	—	Used	ℝ	varies
Commitment to Serve	CMTS_THETA_SCR_QY	Personality	—	Used	ℝ	varies
Cooperation	COOPR_THETA_SCR_QY	Personality	—	Used	ℝ	varies
Courage	COUR_THETA_SCR_QY	Personality	—	Used	ℝ	varies
Dominance	DOMNC_THETA_SCR_QY	Personality	—	Used	ℝ	varies
Even Tempered	EVTMP_THETA_SCR_QY	Personality	—	Used	ℝ	varies
Intellectual Efficiency	INTLL__EFC_THETA_SCR_QY	Personality	—	Used	ℝ	varies
Non-Delinquency	NON_DLNQY_THETA_SCR_QY	Personality	—	Used	ℝ	varies
Optimism	OPTMSM_THETA_SCR_QY	Personality	—	Used	ℝ	varies
Order	ORD_THETA_SCR_QY	Personality	—	Used	ℝ	varies
Physical Condition	PHY_COND_THETA_SCR_QY	Personality	—	Used	ℝ	varies
Responsibility	RSBY_THETA_SCR_QY	Personality	—	Used	ℝ	varies
Sociability	SCBLTY_THETA_SCR_QY	Personality	—	Used	ℝ	varies
Self-Control	SELF_CTRL_THETA_SCR_QY	Personality	—	Used	ℝ	varies
Situational Awareness	SITNL_AWRNS_THETA_SCR_QY	Personality	—	Used	ℝ	varies
Selflessness	SLFNS_THETA_SCR_QY	Personality	—	Used	ℝ	varies
Team Orientation	TEAM_ORNTN_THETA_SCR_QY	Personality	—	Used	ℝ	varies
Tolerance	TOL_THETA_SCR_QY	Personality	—	Used	ℝ	varies
Outcome Variables						
AUDIT-C Score	AUDITC_TOTALSCORE_MEAN	Alcohol Misuse or Abuse	Counterproductive	Calculated	0–12	3
APFT Score	APFT_TOTALSCORE_MEAN	Physical Fitness	Task	Calculated	0–300	3
Award Count	award_count	Merit	Contextual	Derived	0– ℝ	—
Bad Paper Count	badpaper.overall	Misconduct	Counterproductive	Calculated	0– ℝ	3
Speed of Promotion	SOP.RANK_HIGH.STDZ2	Rank Achievement Time	General	Calculated	ℝ	—
First-Term Attrition	ATTRIT_FIRST_TERM	Separation from Army	Outcome	Calculated	2	—
Character of Service	CHAR_SVC_CD2	Terms of Separation	Outcome	Derived	2	—

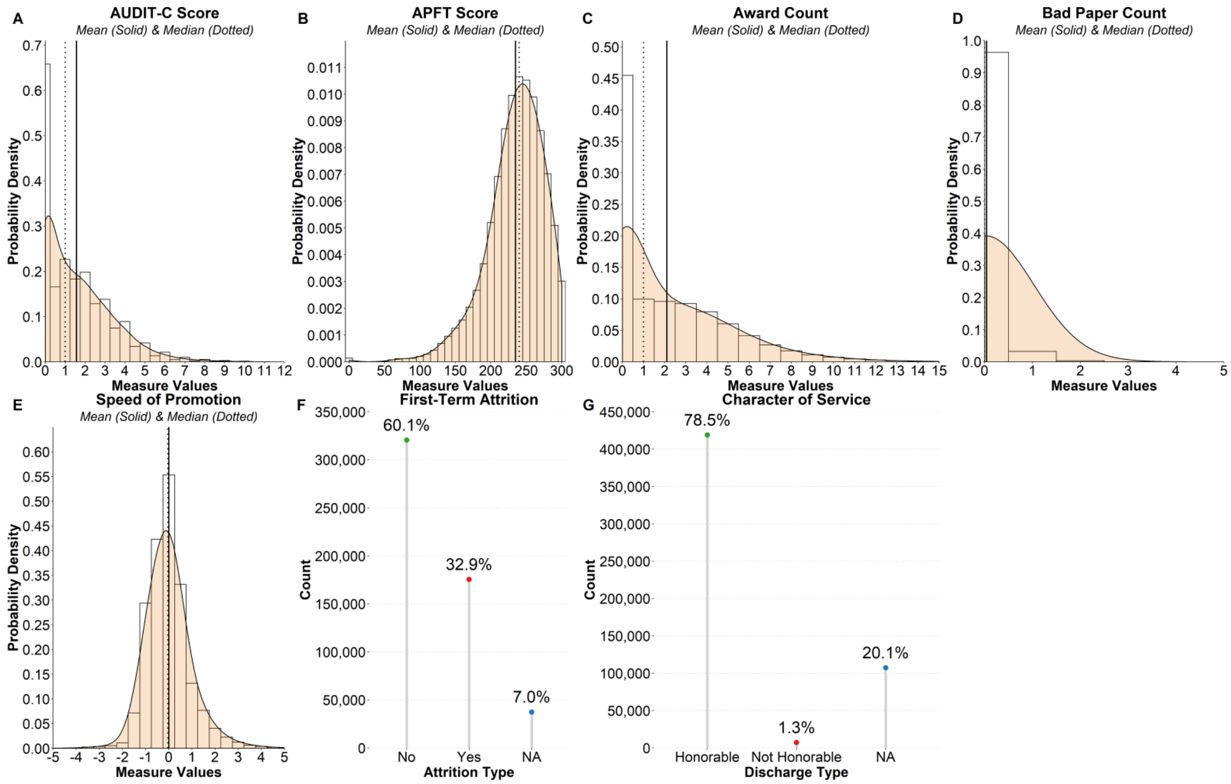
Note. ℝ = any real number. Number of items varies by GAT version (first number = GAT 1.0, second number = GAT 2.0).

Outcomes

The descriptive statistics for outcome variables are also summarized in Exhibit 12. Overall, values (e.g., min/max, means, *SD*s) were consistent with the normal range of the predictor variables. Measures of skewness did not exceed an absolute value of three for most variables except for Bad Paper Count (skewness = 6.34), which is likely higher due to the fact that it is a count variable with most values falling at zero. Measures of (excess) kurtosis did not exceed an absolute value of seven with the exception of Bad Paper Count (kurtosis = 49.20) and Speed of Promotion (kurtosis = 30.99). Thus, most measures of skewness and kurtosis for predictor variables were within acceptable limits (±3 for skewness, ±7 for excess kurtosis; see Hair Jr et al., 2010; Kline, 2011). The distribution pattern for each outcome variable can be found in Exhibit 11. Patterns observed appear to be approximately normal for most variables except for Bad Paper Count. It also appears that receiving the status of a 'not honorable' discharge from the Army is not an event with a high base rate (at 1.3%). Similarly, the 20.1% missing rate for the Character of Service outcome most likely reflects Soldiers who had not discharged from the Army at the time of this research and, therefore, did not have a Character of Service code. Lastly, a correlation matrix of the outcome variables is presented in Exhibit 14. The notable patterns observed within the zero-order correlations (Pearson and point-biserial correlations) between outcome variables were a negative relationship between Award Count and First-Term Attrition, a positive relationship between Award Count and APFT Score, and a positive relationship between Bad Paper Count and Character of Service.

Exhibit 12.

Distributions of Outcome Variables Related to Performance



Note. Solid lines indicate the mean of the sample, and the dotted lines represent the median of sample. Speed of Promotion measure values in months.

Exhibit 13.

Summary of Descriptive Statistics for Predictor Variables

Predictor Variables	Expected Values	*n*	% Missing	# Unique Values	Min	Max	Median	Mean	95% CI_M	*SD*	Skewness	Kurtosis (Excess)	K^2 Normality Test	ω_{Total}
Accession														
Rank Group	2 Levels	533,317	0.00	2	—	—	—	—	—	—	—	—		—
MOS Type	6 Levels	533,226	0.13	6	—	—	—	—	—	—	—	—		—
Soldier Sex	2 Levels	533,317	0.00	2	—	—	—	—	—	—	—	—		—
Soldier Race	6 Levels	533,317	0.00	6	—	—	—	—	—	—	—	—		—
Age at Accession	17–$\mathbb{R}^+$	533,317	0.00	40,740	14.05	68.56	20.61	22.01	[21.99, 22.02]	4.24	2.09	6.00	231,873.41*	—
AFQT Score	1–99	507,927	4.76	99	1	99	57.00	59.69	[59.64, 59.74]	19.44	0.26	−1.01	127,039.06*	—
PULHES Score	1–4	497,529	6.71	16	1	3	1.00	1.07	[1.07, 1.07]	0.13	3.79	33.27	422,639.21*	—
GAT														
Adaptability	1–5	349,464	34.47	13	1	5	4.00	4.02	[4.02, 4.02]	0.77	−0.64	0.15	20,570.29*	0.72
Active Coping	1–5	349,455	34.48	21	1	5	3.80	3.82	[3.82, 3.83]	0.75	−0.50	0.23	13,490.86*	0.80
Passive Coping	1–5	349,455	34.48	13	1	5	2.67	2.60	[2.60, 2.60]	0.93	0.29	−0.30	6,602.47*	0.72
Character	0–10	349,454	34.48	378	0	10	8.17	7.95	[7.94, 7.95]	1.49	−1.14	2.20	67,776.55*	0.96
Catastrophizing	1–5	349,463	34.47	38	1	5	1.86	1.96	[1.96, 1.97]	0.82	0.99	0.91	46,422.99*	0.81
Depression	1–5	349,459	34.47	41	1	5	1.30	1.63	[1.63, 1.63]	0.78	1.69	2.75	106,737.90*	0.93
Optimism	1–5	336,509	36.90	17	1	5	3.75	3.67	[3.66, 3.67]	0.79	−0.31	−0.08	5,186.96*	0.75
Positive Affect	1–5	336,508	36.90	74	1	5	3.90	3.80	[3.80, 3.80]	0.81	−0.52	−0.02	13,690.78*	0.93
Negative Affect	1–5	336,508	36.90	91	1	5	2.09	2.18	[2.18, 2.18]	0.72	0.63	0.44	20,848.87*	0.89
Loneliness	1–5	349,463	34.47	13	1	5	2.00	2.25	[2.24, 2.25]	0.87	0.51	−0.22	14,650.40*	0.81
Organizational Trust	1–5	349,463	34.47	33	1	5	4.00	3.92	[3.92, 3.92]	0.83	−0.84	0.80	36,471.71*	0.86
Work Engagement	1–5	224,437	57.92	17	1	5	4.00	3.76	[3.76, 3.77]	0.95	−0.68	0.03	14,185.87*	0.83
Life Meaning	1–5	349,463	34.48	33	1	5	4.20	4.00	[4.00, 4.00]	0.87	−0.92	0.50	38,735.77*	0.83
TAPAS														
Achievement	$\mathbb{R}$	181,121	66.04	3,108	−2.02	2.36	0.22	0.22	[0.22, 0.22]	0.49	0.02	0.26	401.31*	—
Adjustment	$\mathbb{R}$	124,226	76.71	428	−2.08	2.62	0.05	0.07	[0.07, 0.07]	0.47	0.40	1.17	5,998.60*	—
Adaptation	$\mathbb{R}$	181,121	66.04	6,089	−4.50	4.43	0.09	0.10	[0.09, 0.10]	0.99	0.01	0.02	5.86	—
Adventure	$\mathbb{R}$	29,855	94.40	386	−2.09	2.16	−0.24	−0.24	[−0.25, −0.23]	0.59	0.08	0.06	39.51*	—
Attention Seeking	$\mathbb{R}$	143,243	73.14	3,081	−2.43	2.38	−0.30	−0.30	[−0.30, −0.30]	0.57	0.12	0.43	1,114.28*	—
Commitment to Serve	$\mathbb{R}$	67,915	87.27	2,666	−1.74	1.95	0.24	0.22	[0.22, 0.23]	0.52	−0.30	0.11	1,033.09*	—
Cooperation	$\mathbb{R}$	113,502	78.72	3,013	−2.19	3.13	0.05	0.10	[0.09, 0.10]	0.49	0.79	1.65	13,555.99*	—
Courage	$\mathbb{R}$	48,602	90.89	2,263	−1.72	1.74	0.16	0.17	[0.16, 0.17]	0.54	−0.17	−0.37	646.80*	—
Dominance	$\mathbb{R}$	181,121	66.04	3,340	−2.17	2.63	0.24	0.24	[0.24, 0.24]	0.55	−0.17	0.53	2,253.61*	—
Even Tempered	$\mathbb{R}$	181,121	66.04	3,600	−2.34	3.70	0.32	0.34	[0.34, 0.34]	0.52	0.35	1.59	9,914.90*	—
Intellectual Efficiency	$\mathbb{R}$	181,121	66.04	3,273	−2.54	2.68	0.04	0.04	[0.03, 0.04]	0.54	0.01	0.29	495.23*	—
Non−Delinquency	$\mathbb{R}$	143,061	73.18	2,851	−2.30	2.55	0.17	0.17	[0.16, 0.17]	0.51	−0.07	0.44	901.89*	—
Optimism	$\mathbb{R}$	181,121	66.04	2,976	−2.30	3.32	0.25	0.26	[0.25, 0.26]	0.44	0.15	1.10	4,637.87*	—
Order	$\mathbb{R}$	151,354	71.62	3,449	−2.64	2.14	−0.31	−0.33	[−0.33, −0.32]	0.55	−0.09	0.21	444.58*	—
Physical Condition	$\mathbb{R}$	181,121	66.04	3,617	−2.60	2.42	0.11	0.14	[0.14, 0.15]	0.58	0.20	0.51	2,412.68*	—
Responsibility	$\mathbb{R}$	48,810	90.85	2,604	−1.22	2.69	0.29	0.37	[0.37, 0.37]	0.48	0.73	0.59	3,921.85*	—
Sociability	$\mathbb{R}$	151,266	71.64	3,568	−2.37	2.76	−0.13	−0.14	[−0.15, −0.14]	0.59	−0.01	0.12	81.63*	—
Self−Control	$\mathbb{R}$	94,371	82.30	405	−2.03	2.23	−0.10	−0.07	[−0.08, −0.07]	0.52	0.17	0.09	492.92*	—
Situational Awareness	$\mathbb{R}$	48,690	90.87	2,598	−1.97	2.24	0.05	0.05	[0.05, 0.06]	0.49	0.00	0.11	21.74*	—
Selflessness	$\mathbb{R}$	151,354	71.62	2,941	−2.50	2.67	−0.01	−0.02	[−0.02, −0.02]	0.46	−0.01	0.43	792.25*	—
Team Orientation	$\mathbb{R}$	48,784	90.85	2,769	−1.96	2.41	0.00	0.01	[0.00, 0.01]	0.50	0.07	0.53	401.25*	—
Tolerance	$\mathbb{R}$	151,266	71.64	3,336	−2.51	2.65	0.00	−0.03	[−0.03, −0.03]	0.56	−0.13	0.24	714.89*	—

Note. $\mathbb{R}$ = any real number; * = $p < .05$.

Exhibit 14.

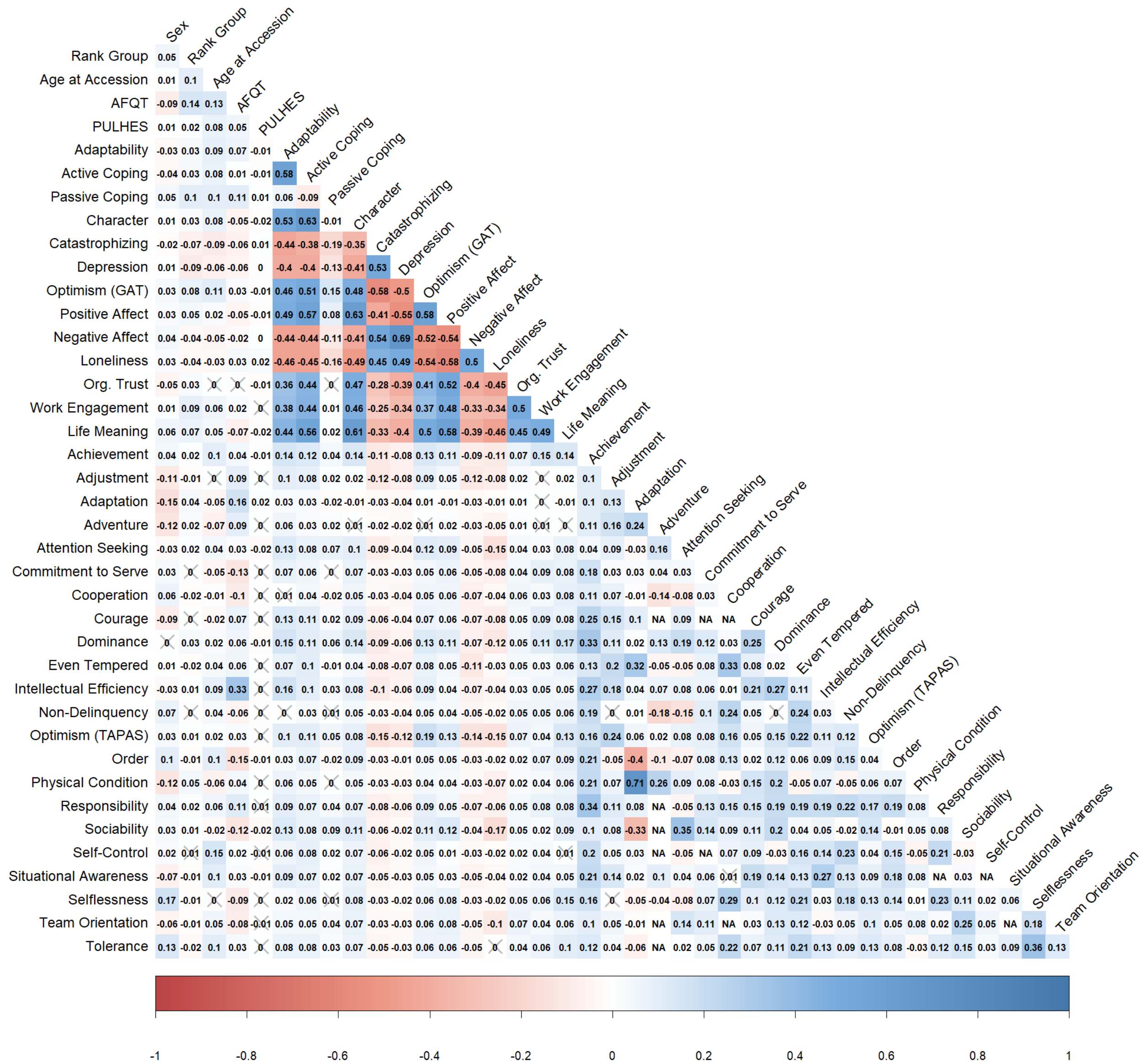
*Summary of Descriptive Statistics f***or Outcome Variables**

Outcome Variables	Expected Values	n	% Missing	# Unique Values	Min	Max	Median	Mean	95% CI_M	SD	Skewness	Kurtosis (Excess)	K^2 Normality Test	ω_{Total}
AUDIT−C Score	0–12	453,390	14.99	228	0	12	1.00	1.57	[1.56, 1.57]	1.67	1.26	1.80	93,076.07*	—
APFT Score	0–300	404,558	24.14	11,902	0	300	239.80	234.79	[234.66, 234.91]	40.78	−1.08	2.37	76,388.13*	—
Award Count	0– $\mathbb{R}^+$	533,317	0.00	44	0	54	1.00	2.10	[2.10, 2.11]	2.73	1.70	4.82	186,538.29*	—
Bad Paper Count	0– $\mathbb{R}^+$	533,317	0.00	6	0	5	0.00	0.04	[0.04, 0.04]	0.22	6.34	49.20	622,420.68*	—
Speed of Promotion	$\mathbb{R}^-$– $\mathbb{R}^+$	462,287	13.33	5,077	−6.77	33.21	−0.06	0.00	[0.00, 0.00]	1.00	2.44	30.99	301,559.23*	—
First−Term Attrition	2 Levels	495,932	7.01	2	0	1	—	—	—	—	—	—	—	—
Character of Service	2 Levels	426,066	20.11	2	0	1	—	—	—	—	—	—	—	—

Note. $\mathbb{R}^+$ = any real positive number, $\mathbb{R}^-$ = any real negative number; * = $p < .05$.

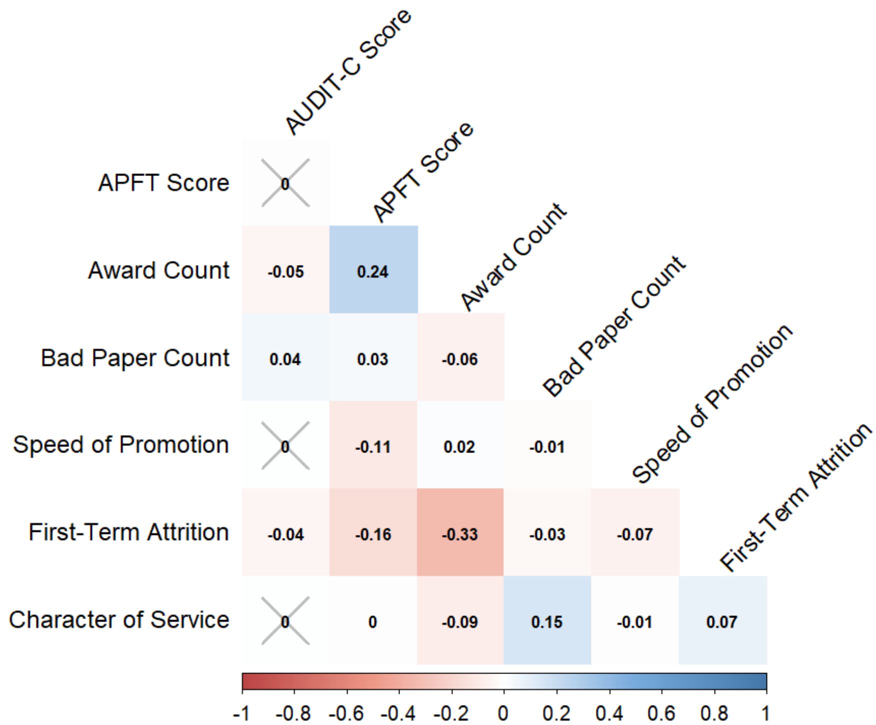
Exhibit 15.

Zero-Order **Correlations between Predictor Variables**



Note. '×' marks represent correlations that were not significant at the $p < .05$ level. Correlations represent zero-order Pearson coefficients. For correlations with categorical variables, a point-biserial coefficient is calculated.

Exhibit 16.
Zero-Order Correlations between Outcome Variables



Note. '×' marks represent correlations that were not significant at the $p < .05$ level. Correlations represent zero-order Pearson coefficients. For correlations with categorical variables, a point-biserial coefficient is calculated.

Simple Relationship Analysis

The aim of the simple relationship analysis was to understand the size and direction of the relationships between predictors and outcomes related to performance. Specifically, a simple analysis examines the relationship of a single predictor with a single outcome variable (e.g., zero-order correlation, simple linear regression, *t*-test, one-way ANOVA). Given that there were 42 predictors identified and seven outcome variables of interest, this meant that there would be 294 simple models analyzed. Analyses were conducted on an overall sample of enlisted Soldiers and officers as well as subsamples examining only enlisted Soldiers and officers separately to provide points of comparisons across these groups.

Family-wise corrections for probability values were also made to account for the multiple tests. The family-wise error rate refers to the probability of making one or more type I errors when conducting multiple hypotheses tests of the same type, often because the hypotheses resting on tests that shame some amount of error or because probability of observing a statistically significant result increases as the total number of statistical test increases. In our study, a Bonferroni correction was applied. This correction compensates for family-wise error rates by testing the individual hypothesis with a significance level set to α / m, where α is the overall alpha level (e.g., 0.05) and m is the total number of hypotheses. For example, in Exhibit 20, displaying the results of the AUDIT-C analysis, we see that the Bonferroni corrected p-value is equal to $\alpha/882 = 0.000057$, as our overall α is $p < 0.05$.

Assumption checks were used to help determine appropriate modeling procedures (for an example of this process, see the “Benchmarking Crosswalks of Army data from Multiple Sources” report; Ratcliff et al., unpublished). In general, most of the continuous outcome and predictor variables followed a relatively normal distribution upon examining descriptive statistics and graphical representations of normality and linearity. For an exception, see Bad Paper Count below which warranted the use of a generalized linear model using a log link function. Statistical tests were then chosen on the basis of predictor and outcome variables and included correlations, simple linear regressions, generalized linear models (with log and logit linkages), *t*-tests and one-way analysis of variance (ANOVA) for categorical predictors (i.e., Rank Group, MOS Type, Soldier Sex, Soldier Race). For categorical predictors, models were examined using traditional parametric procedures (i.e., Student’s *t*-test, Fisher’s *F*-test) alongside models robust to heterogeneity and non-parametric models. Results between traditional parametric tests and alternative models were nearly identical in terms of direction and magnitude. Given this finding, we report the traditional parametric tests for ease of interpretation. This approach is supported given the large sample sizes and the fact that traditional models like the ANOVA are very robust to assumption violations like departures from normality (see Blanca et al., 2017; Schmider et al., 2010).

Lastly, descriptive statistics (e.g., means, standard deviations, counts) are reported to provide further information related to the direction of relationships. For continuous predictors, a ‘(+)’ indicates a positive relationship and ‘(−)’ indicates a negative relationship with the outcome variable. For each outcome variable, standardized effect sizes are calculated to provide comparisons amongst the predictor variables and the sampling groups (i.e., enlisted + officer, enlisted only, officer only).

AUDIT-C Score

To examine the simple relationships of Audit-C Scores, a series of ANOVAs (Analysis of Variance) were conducted for categorical predictors and simple linear regressions were conducted for continuous predictors. An ANOVA is a statistical test that compares the means of two or more groups to determine if there is a significant difference between them. A summary of standardized effect sizes for each predictor is presented in Exhibit 18.

Categorical Predictors. For a summary of results, see Exhibit 16. Rank Group, Soldier Sex, and Soldier Race all had small effects. For Rank Group, officers ($M = 2.17$, $SD = 1.48$) showed higher levels of AUDIT-C Scores than enlisted Soldiers ($M = 1.52$, $SD = 1.68$), $h^2 = 0.01$, 90% Confidence Interval (CI) [0.01, 0.01]. For Soldier Sex, male Soldiers ($M = 1.66$, $SD = 1.73$) showed higher levels than female Soldiers ($M = 1.05$, $SD = 1.18$), $h^2 = 0.02$, 90% CI [0.02, 0.02]. For Soldier Race, Soldiers of Mixed Race/Other ($M = 1.93$, $SD = 1.51$) showed the highest levels while Native

Hawaiians/Pacific Islanders (*M* = 1.22, *SD* = 1.59). Similar patterns were found when looking at only enlisted Soldiers and officers. However, the effects for officers seemed to be of greater strength and officers showed an MOS Type effect where Combat Arms Soldiers showed the highest levels (*M* = 2.47, *SD* = 1.55) and Combat Service Support Soldiers showed the lowest (*M* = 1.38, *SD* = 1.55).

Continuous Predictors. For a summary of the results, see Exhibit 17. Several continuous predictors showed small effects with AUDIT-C Scores. Positive relationships included Age at Accession (+), AFQT Score (+), and Attention Seeking (+). Negative relationships included Character (−), Positive Affect (−), Life Meaning (−), Cooperation (−), Order (−), and Selflessness (−). Of these, Life Meaning had the strongest relationship with an r^2 value of 0.02. Similar patterns were found when looking at only enlisted Soldiers. For officers, some patterns reversed in the case of Age at Accession (−) along with new patterns such as Depression (+), Adventure (+), Optimism (GAT; −), and Team Orientation (−) to name a few.

Exhibit 17.
Categorical Predictors with AUDIT-C Score by Sample Type

Predictors	*n*	*F*	h^2	90 % CI	*p*	Group *M*s (*SD*s)
Enlisted + Officer						
Rank Group	453,390	4,560.30	**0.01**	[0.01, 0.01]	< .001*	*Enlisted*: 1.52 (1.68) *Officer*: 2.17 (1.48)
MOS Type	453,305	375.27	0.00	[0.00, 0.00]	< .001*	*Combat Arms*: 1.67 (1.78) *Combat Support*: 1.50 (1.55) *Combat Service Support*: 1.42 (1.56) *Special Service*: 1.57 (1.60) *Operations*: 1.90 (1.63) *Unknown*: 1.48 (1.52)
Soldier Sex	453,390	7,786.30	**0.02**	[0.02, 0.02]	< .001*	*Male*: 1.66 (1.73) *Female*: 1.05 (1.18)
Soldier Race	453,390	867.39	**0.01**	[0.01, 0.01]	< .001*	*White*: 1.66 (1.72) *Black*: 1.27 (1.50) *Asian*: 1.44 (1.51) *American Indian/Alaskan Native*: 1.50 (1.72) *Native Hawaiian/Pacific Islander*: 1.22 (1.59) *Mixed Race/Other*: 1.93 (1.51)
Enlisted						
Rank Group	—	—	—	—	—	—
MOS Type	421,507	306.56	0.00	[0.00, 0.00]	< .001*	*Combat Arms*: 1.61 (1.78) *Combat Support*: 1.45 (1.54) *Combat Service Support*: 1.38 (1.55) *Special Service*: 1.54 (1.68) *Operations*: 1.92 (1.65) *Unknown*: 1.47 (1.52)
Soldier Sex	421,507	7,492.63	**0.02**	[0.02, 0.02]	< .001*	*Male*: 1.61 (1.73) *Female*: 0.98 (1.18)
Soldier Race	421,507	648.28	**0.01**	[0.01, 0.01]	< .001*	*White*: 1.61 (1.72) *Black*: 1.26 (1.50) *Asian*: 1.40 (1.52) *American Indian/Alaskan Native*: 1.48 (1.73) *Native Hawaiian/Pacific Islander*: 1.20 (1.60) *Mixed Race/Other*: 1.86 (1.66)
Officer						
Rank Group	—	—	—	—	—	—
MOS Type	31,798	319.29	**0.05**	[0.04, 0.05]	< .001*	*Combat Arms*: 2.47 (1.55) *Combat Support*: 2.17 (1.39) *Combat Service Support*: 2.25 (1.49) *Special Service*: 1.71 (1.24) *Operations*: 1.68 (1.28) *Unknown*: 2.08 (1.49)
Soldier Sex	31,883	1,386.07	**0.04**	[0.04, 0.05]	< .001*	*Male*: 2.34 (1.54) *Female*: 1.62 (1.06)
Soldier Race	31,883	130.79	**0.02**	[0.02, 0.02]	< .001*	*White*: 2.28 (1.49) *Black*: 1.61 (1.31) *Asian*: 1.80 (1.36) *American Indian/Alaskan Native*: 1.99 (1.36) *Native Hawaiian/Pacific Islander*: 1.69 (1.31) *Mixed Race/Other*: 1.96 (1.42)

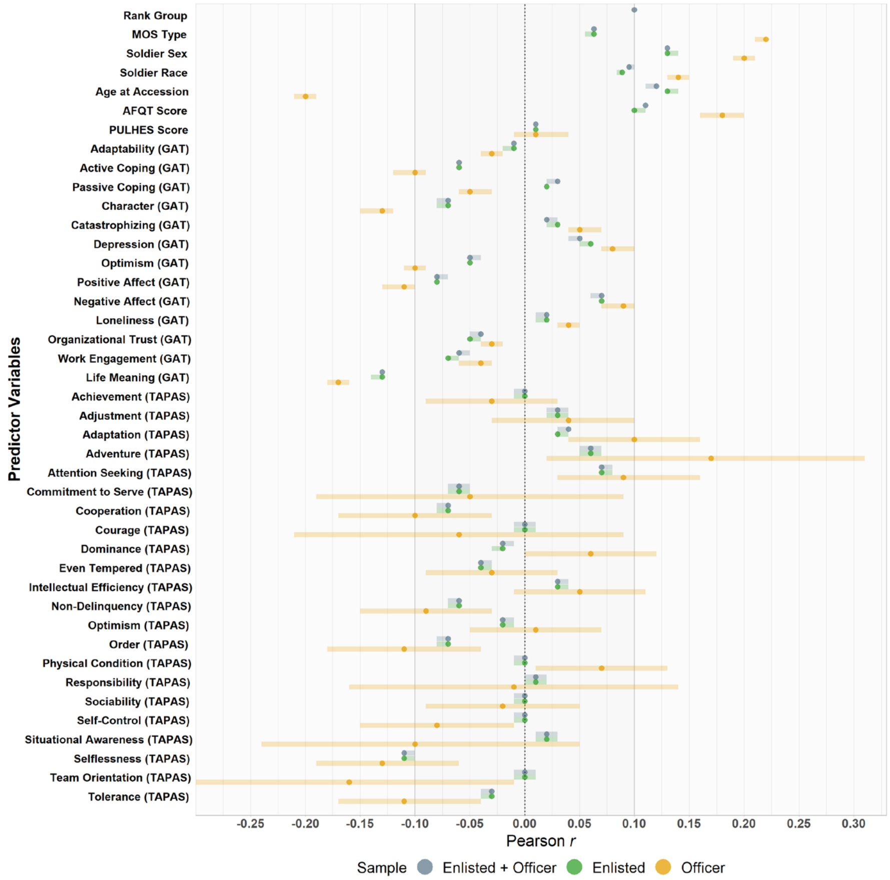
Note. AUDIT-C = The Alcohol Use Disorders Identification Test-Concise, *n* = sample size, *F* = *F*-test coefficient, h^2 = eta squared, CI = confidence interval of effect size, *p* = p-value, $*p < .05$ after applying family-wise Bonferroni correction (α/882 = 0.000057); effect size values in bold represent effect sizes greater than trivial according to Cohen (1988).

Exhibit 18.
Continuous Predictors with AUDIT-C Score by Sample Type

Predictors	Enlisted + Officer						Enlisted						Officer					
	n	*b*	95% CI	*β*	r^2	*p*	*n*	*b*	95% CI	*β*	r^2	*p*	*n*	*b*	95% CI	*β*	r^2	*p*
Accession																		
Age at Accession	453,216	0.05	[0.04, 0.05]	0.12	**0.01**	< .001*	421,507	0.05	[0.05, 0.05]	0.13	**0.02**	< .001*	31,883	−0.07	[−0.06, −0.06]	−0.20	**0.04**	< .001*
AFQT Score	430,268	0.01	[0.01, 0.01]	0.11	**0.01**	< .001*	421,142	0.01	[0.01, 0.01]	0.10	**0.01**	< .001*	9,126	0.02	[0.01, 0.02]	0.18	**0.03**	< .001*
PULHES Score	421,003	0.14	[0.10, 0.18]	0.01	0.00	< .001*	421,507	0.14	[0.10, 0.18]	0.01	0.00	< .001*	7,943	0.11	[−0.06, 0.27]	0.01	0.00	.212
GAT																		
Adaptability	315,008	−0.03	[−0.03, −0.02]	−0.01	0.00	< .001*	291,305	−0.03	[−0.04, −0.02]	−0.01	0.00	< .001*	23,701	−0.06	[−0.09, −0.03]	−0.03	0.00	< .001*
Active Coping	314,999	−0.14	[−0.14, −0.13]	−0.06	0.00	< .001*	291,296	−0.14	[−0.14, −0.13]	−0.06	0.00	< .001*	23,701	−0.25	[−0.28, −0.22]	−0.10	**0.01**	< .001*
Passive Coping	314,999	0.05	[0.04, 0.06]	0.03	0.00	< .001*	291,296	0.04	[0.03, 0.04]	0.02	0.00	< .001*	23,701	−0.08	[−0.10, −0.06]	−0.05	0.00	< .001*
Character	314,999	−0.08	[−0.09, −0.08]	−0.07	**0.01**	< .001*	291,299	−0.08	[−0.09, −0.08]	−0.07	**0.01**	< .001*	23,698	−0.17	[−0.18, −0.15]	−0.13	**0.02**	< .001*
Catastrophizing	315,008	0.04	[0.04, 0.05]	0.02	0.00	< .001*	291,306	0.06	[0.05, 0.06]	0.03	0.00	< .001*	23,700	0.11	[0.09, 0.14]	0.05	0.00	< .001*
Depression	315,004	0.10	[0.09, 0.11]	0.05	0.00	< .001*	291,301	0.12	[0.11, 0.13]	0.06	0.00	< .001*	23,701	0.23	[0.20, 0.27]	0.08	**0.01**	< .001*
Optimism	283,229	−0.10	[−0.10, −0.09]	−0.05	0.00	< .001*	280,309	−0.11	[−0.11, −0.10]	−0.05	0.00	< .001*	22,918	−0.22	[−0.25, −0.19]	−0.10	**0.01**	< .001*
Positive Affect	303,228	−0.16	[−0.17, −0.15]	−0.08	**0.01**	< .001*	280,307	−0.17	[−0.17, −0.16]	−0.08	**0.01**	< .001*	22,919	−0.26	[−0.29, −0.23]	−0.11	**0.01**	< .001*
Negative Affect	303,228	0.16	[0.15, 0.17]	0.07	0.00	< .001*	280,307	0.16	[0.15, 0.17]	0.07	0.00	< .001*	22,919	0.23	[0.19, 0.26]	0.09	**0.01**	< .001*
Loneliness	315,008	0.03	[0.02, 0.04]	0.02	0.00	< .001*	291,306	0.03	[0.03, 0.04]	0.02	0.00	< .001*	23,700	0.08	[0.05, 0.11]	0.04	0.00	< .001*
Organizational Trust	315,007	−0.09	[−0.09, −0.08]	−0.04	0.00	< .001*	291,306	−0.10	[−0.10, −0.09]	−0.05	0.00	< .001*	23,699	−0.06	[−0.09, −0.03]	−0.03	0.00	< .001*
Work Engagement	203,324	−0.11	[−0.11, −0.10]	−0.06	0.00	< .001*	187,209	−0.12	[−0.13, −0.11]	−0.07	0.00	< .001*	16,111	−0.08	[−0.11, −0.05]	−0.04	0.00	< .001*
Life Meaning	314,999	−0.25	[−0.25, −0.24]	−0.13	**0.02**	< .001*	291,297	−0.26	[−0.26, −0.25]	−0.13	**0.02**	< .001*	23,700	−0.33	[−0.35, −0.30]	−0.17	**0.03**	< .001*
TAPAS																		
Achievement	150,875	−0.01	[−0.03, 0.01]	0.00	0.00	.221	149,857	−0.01	[−0.03, 0.01]	0.00	0.00	.184	1,016	−0.08	[−0.26, 0.09]	−0.03	0.00	.353
Adjustment	106,896	0.10	[0.08, 0.12]	0.03	0.00	< .001*	105,908	0.10	[0.08, 0.12]	0.03	0.00	< .001*	989	0.10	[−0.07, 0.26]	0.04	0.00	.250
Adaptation	149,114	0.06	[0.05, 0.06]	0.04	0.00	< .001*	149,857	0.05	[0.05, 0.06]	0.03	0.00	< .001*	1,016	0.13	[0.06, 0.24]	0.10	**0.01**	.002
Adventure	25,507	0.16	[0.13, 0.19]	0.06	0.00	< .001*	25,334	0.15	[0.12, 0.18]	0.06	0.00	< .001*	171	0.46	[0.06, 0.85]	0.17	**0.03**	.025
Attention Seeking	121,659	0.20	[0.18, 0.21]	0.07	**0.01**	< .001*	120,659	0.20	[0.18, 0.21]	0.07	**0.01**	< .001*	998	0.23	[0.08, 0.37]	0.09	**0.01**	.003
Commitment to Serve	54,960	−0.17	[−0.20, −0.15]	−0.06	0.00	< .001*	54,769	−0.17	[−0.20, −0.15]	−0.06	0.00	< .001*	189	−0.13	[−0.49, 0.24]	−0.05	0.00	.502
Cooperation	96,133	−0.23	[−0.25, −0.21]	−0.07	**0.01**	< .001*	95,324	−0.23	[−0.24, −0.21]	−0.07	**0.01**	< .001*	834	−0.32	[−0.54, −0.10]	−0.10	**0.01**	.004
Courage	39,957	0.01	[−0.02, 0.04]	0.00	0.00	.485	39,782	0.01	[−0.02, 0.04]	0.00	0.00	.461	173	−0.14	[−0.52, 0.23]	−0.06	0.00	.452
Dominance	150,871	−0.05	[−0.07, −0.04]	−0.02	0.00	< .001*	149,857	−0.06	[−0.07, −0.04]	−0.02	0.00	< .001*	1,016	0.15	[−0.01, 0.30]	0.06	0.00	.071
Even Tempered	150,816	−0.11	[−0.12, −0.09]	−0.04	0.00	< .001*	149,857	0.10	[−0.12, −0.09]	−0.04	0.00	< .001*	1,016	−0.10	[−0.28, 0.09]	−0.03	0.00	.301
Intellectual Efficiency	150,874	0.09	[0.07, 0.10]	0.03	0.00	< .001*	149,857	0.09	[0.07, 0.10]	0.03	0.00	< .001*	1,016	0.12	[−0.04, 0.27]	0.05	0.00	.130
Non-Delinquency	121,420	−0.19	[−0.21, −0.18]	−0.06	0.00	< .001*	120,422	−0.19	[−0.21, −0.18]	−0.06	0.00	< .001*	998	−0.24	[−0.43, −0.08]	−0.09	**0.01**	.005
Optimism	150,850	−0.07	[−0.08, −0.05]	−0.02	0.00	< .001*	149,857	−0.07	[−0.08, −0.05]	−0.02	0.00	< .001*	1,016	0.04	[−0.15, 0.21]	0.01	0.00	.725
Order	125,441	−0.21	[−0.22, −0.19]	−0.07	**0.01**	< .001*	124,589	−0.21	[−0.22, −0.19]	−0.07	**0.01**	< .001*	852	−0.26	[−0.41, −0.10]	−0.11	**0.01**	.001
Physical Condition	150,873	−0.01	[−0.02, 0.01]	0.00	0.00	.411	149,857	−0.01	[−0.02, 0.00]	0.00	0.00	.147	1,016	0.17	[0.02, 0.31]	0.07	0.00	.027
Responsibility	40,125	0.03	[−0.01, 0.06]	0.01	0.00	.110	39,952	0.02	[−0.01, 0.05]	0.01	0.00	.134	173	−0.02	[−0.44, 0.39]	−0.01	0.00	.918
Sociability	125,368	−0.01	[−0.02, 0.01]	0.00	0.00	.227	124,523	−0.01	[−0.02, 0.01]	0.00	0.00	.216	845	−0.04	[−0.19, 0.11]	−0.02	0.00	.602
Self-Control	81,392	−0.01	[−0.03, 0.01]	0.00	0.00	.164	80,574	−0.01	[−0.03, 0.01]	0.00	0.00	.217	818	−0.20	[−0.38, −0.03]	−0.08	**0.01**	.024
Situational Awareness	40,028	0.05	[0.02, 0.08]	0.02	0.00	.001	39,848	0.06	[0.03, 0.09]	0.02	0.00	< .001	180	−0.32	[−0.80, 0.15]	−0.10	**0.01**	.185
Selflessness	125,441	−0.36	[−0.38, −0.34]	−0.11	**0.01**	< .001*	124,589	−0.36	[−0.38, −0.34]	−0.11	**0.01**	< .001*	852	−0.37	[−0.56, −0.17]	−0.13	**0.02**	< .001
Team Orientation	40,192	−0.01	[−0.04, 0.02]	0.00	0.00	.590	40,019	−0.01	[−0.04, 0.02]	0.00	0.00	.714	173	−0.45	[−0.87, −0.02]	−0.16	**0.02**	.040
Tolerance	125,368	−0.09	[−0.10, −0.07]	−0.03	0.00	< .001*	124,523	−0.09	[−0.10, −0.07]	−0.03	0.00	< .001*	845	−0.25	[−0.40, −0.09]	−0.11	**0.01**	.002

Note. AUDIT-C = The Alcohol Use Disorders Identification Test-Concise, GAT = Global Assessment Tool, TAPAS = Tailor Adaptive Personality Assessment System, *n* = sample size, *b* = unstandardized coefficient, CI = confidence interval, *β* = standardized coefficient (equivalent to *r*), *p* = p-value, $*p < .05$ after applying family-wise Bonferroni correction ($\alpha/882 = 0.000057$); r^2 values in bold represent effect sizes greater than trivial according to Cohen (1988).

Exhibit 19.
Plot of Simple Relationship Strength between Predictor Variables and AUDIT-C Score



Note. Shaded area represents trivial effect size according to Cohen (1988).

APFT Score

To examine the simple relationships of APFT Scores, a series of ANOVAs were conducted for categorical predictors and simple linear regressions were conducted for continuous predictors. A summary of standardized effect sizes for each predictor is presented in Exhibit 21.

Categorical Predictors. For a summary of the results, see Exhibit 19. Rank Group and MOS Type all had small effects. For Rank Group, officers (M = 263.96, SD = 36.55) showed higher levels of APFT Scores than enlisted Soldiers (M = 232.59, SD = 40.24), h2 = 0.04, 90% CI [0.04, 0.04]. For MOS Type, Combat Arms Soldiers showed the highest scores (M = 239.03, SD = 39.43) while Combat Service Support Soldiers showed the lowest scores (M = 228.96, SD = 41.95). Similar patterns were found when looking at only enlisted Soldiers and officers. However, officers also showed a small Soldier Sex effect such that male Soldiers (M = 265.86, SD = 33.49) showed higher scores than female Soldiers (M = 257.87, SD = 44.43), h2 = 0.01, 90% CI [0.01, 0.01].

Continuous Predictors. For a summary of results, see Exhibit 20. Several continuous predictors showed small effects with APFT Scores. Positive relationships included Adaptability (+), Active Coping (+), Character (+), Optimism (GAT; +), Positive Affect (+), Organizational Trust (+), Work Engagement (+), Life Meaning (+), Adaptation (+), Adventure (+), and Physical Condition (+). Negative relationships included Catastrophizing (−), Depression (−), Negative Affect (−), and Loneliness (−). Overall, Physical Condition had the strongest relationship with an r^2 value of 0.05. Similar patterns were found when looking at only enlisted Soldiers. For officers, similar effect sizes were observed for the TAPAS measures but none of the GAT measures reached a small effect size. However, additional patterns were found for officers including a medium effect (r^2 = 0.14) of Age at Accession (−) and small effects of Adventure (+), Dominance (+), Non-Delinquency (−), and Situational Awareness (−).

Exhibit 20.

Categorical Predictors with APFT Score by Sample Type

Predictors	*n*	*F*	h^2	90 % CI	*p*	Group *M*s (*SD*s)
Enlisted + Officer						
Rank Group	404,558	16,200.20	**0.04**	[0.04, 0.04]	< .001*	*Enlisted*: 232.59 (40.24) *Officer*: 263.96 (36.55)
MOS Type	404,502	994.79	**0.01**	[0.01, 0.01]	< .001*	*Combat Arms*: 239.03 (39.43) *Combat Support*: 232.77 (40.61) *Combat Service Support:* 228.96 (41.95) *Special Service*: 230.55 (42.11) *Operations*: 231.91 (46.26) *Unknown*: 238.91 (39.64)
Soldier Sex	404,558	1,103.57	0.00	[0.00, 0.00]	< .001*	*Male*: 235.68 (39.63) *Female*: 229.72 (46.49)
Soldier Race	404,558	229.03	0.00	[0.00, 0.00]	< .001*	*White*: 234.77 (40.92) *Black*: 233.15 (40.60) *Asian*: 239.82 (38.11) *American Indian/Alaskan Native*: 232.08 (40.92) *Native Hawaiian/Pacific Islander*: 232.30 (37.45) *Mixed Race/Other*: 255.18 (40.88)
Enlisted						
Rank Group	—	—	—	—	—	—
MOS Type	376,239	972.02	**0.01**	[0.01, 0.01]	< .001*	*Combat Arms*: 236.57 (38.96) *Combat Support*: 230.05 (39.93) *Combat Service Support*: 227.16 (41.63) *Special Service*: 227.30 (40.57) *Operations*: 232.08 (46.41) *Unknown*: 238.71 (39.57)
Soldier Sex	376,239	1,589.62	0.00	[0.00, 0.00]	< .001*	*Male*: 233.66 (39.18) *Female*: 226.21 (45.54)
Soldier Race	376,239	59.66	0.00	[0.00, 0.00]	< .001*	*White*: 232.30 (40.31) *Black*: 232.59 (40.47) *Asian*: 237.53 (37.58) *American Indian/Alaskan Native*: 230.55 (40.50) *Native Hawaiian/Pacific Islander*: 230.42 (36.79) *Mixed Race/Other*: 234.85 (45.69)
Officer						
Rank Group	—	—	—	—	—	—
MOS Type	28,263	909.25	**0.14**	[0.13, 0.15]	< .001*	*Combat Arms*: 275.40 (26.20) *Combat Support*: 272.80 (27.31) *Combat Service Support*: 265.96 (29.53) *Special Service*: 243.99 (45.57) *Operations*: 229.62 (44.74) *Unknown*: 267.11 (39.24)
Soldier Sex	28,319	247.75	**0.01**	[0.01, 0.01]	< .001*	*Male*: 265.86 (33.49) *Female*: 257.87 (44.43)
Soldier Race	28,319	26.16	0.00	[0.00, 0.01]	< .001*	*White*: 264.81 (36.08) *Black*: 256.00 (42.56) *Asian*: 260.78 (36.56) *American Indian/Alaskan Native*: 265.99 (35.38) *Native Hawaiian/Pacific Islander*: 271.39 (28.76) *Mixed Race/Other*: 265.40 (33.91)

Note. APFT = Army Physical Fitness Test, *n* = sample size, *F* = *F*-test coefficient, h^2 = eta squared, CI = confidence interval of effect size, *p* = p-value, **p* < .05 after applying family-wise Bonferroni correction (α/882 = 0.000057); effect size values in bold represent effect sizes greater than trivial according to Cohen (1988).

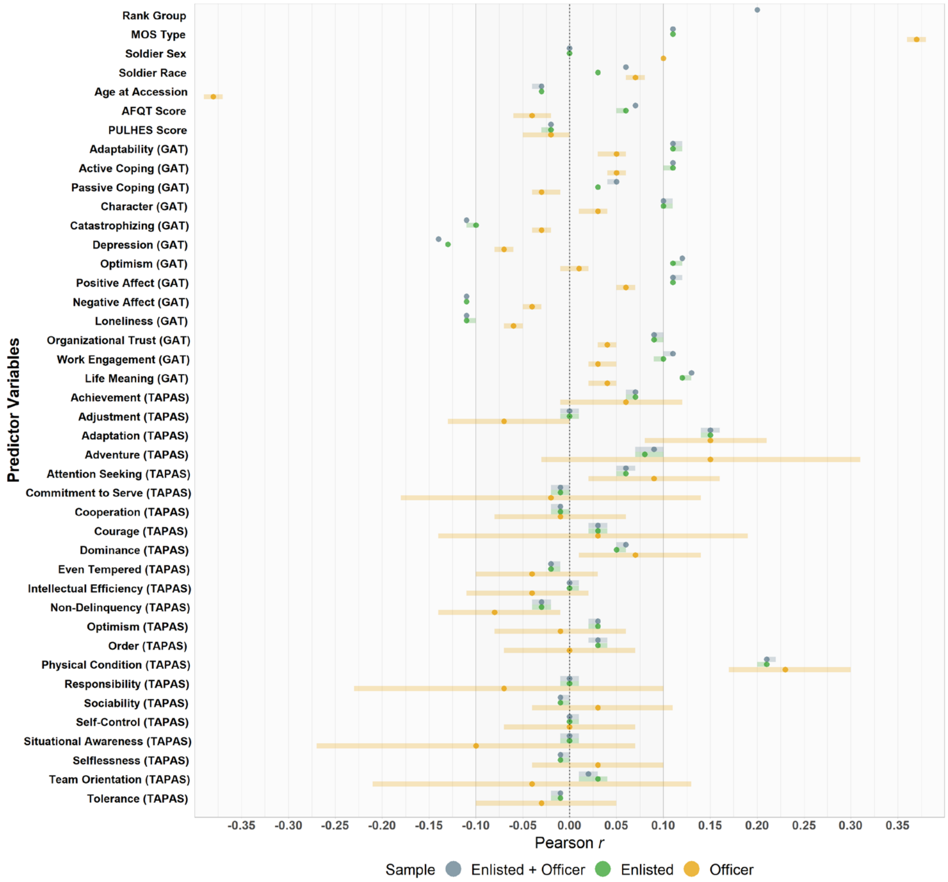
Exhibit 21.
Continuous Predictors with APFT Score by Sample Type

	Enlisted + Officer						*Enlisted*						*Officer*					
Predictors	***n***	***b***	**95% CI**	***β***	***r*²**	***p***	***n***	***b***	**95% CI**	***β***	***R*²**	***p***	***n***	***b***	**95% CI**	***β***	***r*²**	***p***
Accession																		
Age at Accession	404,558	−0.31	[−0.36, −0.30]	−0.03	0.00	< .001*	376,239	−0.28	[−0.32, −0.26]	−0.03	0.00	< .001*	28,319	−2.90	[−2.92, −2.76]	−0.38	**0.14**	< .001*
AFQT Score	404,558	0.15	[0.14, 0.15]	0.07	0.00	< .001*	375,988	0.12	[0.11, 0.12]	0.06	0.00	< .001*	8,461	−0.08	[−0.12, −0.04]	−0.04	0.00	< .001
PULHES Score	375,757	−6.54	[−7.54, −5.54]	−0.02	0.00	< .001*	368,380	−7.17	[−8.19, −6.15]	−0.02	0.00	< .001*	7,377	−4.35	[−8.66, −0.03]	−0.02	0.00	.048
GAT																		
Adaptability	283,410	6.16	[5.96, 6.35]	0.11	**0.01**	< .001*	262,171	6.01	[5.81, 6.21]	0.11	**0.01**	< .001*	21,239	2.54	[1.81, 3.26]	0.05	0.00	< .001*
Active Coping	283,404	5.95	[5.75, 6.15]	0.11	**0.01**	< .001*	262,165	5.77	[5.57, 5.97]	0.11	**0.01**	< .001*	21,239	2.87	[2.09, 3.66]	0.05	0.00	< .001*
Passive Coping	283,404	2.09	[1.92, 2.25]	0.05	0.00	< .001*	262,165	1.30	[1.14, 1.47]	0.03	0.00	< .001*	21,239	−1.03	[−1.59, −0.48]	−0.03	0.00	< .001
Character	283,400	2.82	[2.72, 2.92]	0.10	**0.01**	< .001*	262,164	2.77	[2.67, 2.87]	0.10	**0.01**	< .001*	21,236	0.86	[0.45, 1.27]	0.03	0.00	< .001*
Catastrophizing	283,409	−5.42	[−5.60, −5.24]	−0.11	**0.01**	< .001*	262,171	−4.97	[−5.16, −4.79]	−0.10	**0.01**	< .001*	21,238	−1.60	[−2.31, −0.89]	−0.03	0.00	< .001*
Depression	283,407	−7.45	[−7.64, −7.26]	−0.14	**0.02**	< .001*	262,168	−6.61	[−6.80, −6.41]	−0.13	**0.02**	< .001*	21,239	−4.74	[−5.66, −3.82]	−0.07	0.00	< .001*
Optimism	273,546	6.26	[6.07, 6.45]	0.12	**0.02**	< .001*	252,927	5.78	[5.59, 5.98]	0.11	**0.01**	< .001*	20,619	0.35	[−0.36, 1.06]	0.01	0.00	.335
Positive Affect	273,547	5.80	[5.61, 5.99]	0.11	**0.01**	< .001*	252,927	5.46	[5.27, 5.65]	0.11	**0.01**	< .001*	20,620	3.31	[2.57, 4.05]	0.06	0.00	< .001*
Negative Affect	273,547	−6.36	[−6.57, −6.14]	−0.11	**0.01**	< .001*	252,927	−6.13	[−6.35, −5.92]	−0.11	**0.01**	< .001*	20,620	−2.60	[−3.47, −1.74]	−0.04	0.00	< .001*
Loneliness	283,409	−5.19	[−5.36, −5.01]	−0.11	**0.01**	< .001*	262,171	−4.96	[−5.13, −4.78]	−0.11	**0.01**	< .001*	21,238	−2.99	[−3.64, −2.34]	−0.06	0.00	< .001*
Organizational Trust	283,409	4.64	[4.46, 4.82]	0.09	**0.01**	< .001*	262,172	4.47	[4.29, 4.66]	0.09	**0.01**	< .001*	21,237	2.18	[1.47, 2.88]	0.04	0.00	< .001*
Work Engagement	188,212	4.94	[4.74, 5.14]	0.11	**0.01**	< .001*	173,049	4.27	[4.06, 4.48]	0.10	**0.01**	< .001*	15,163	1.57	[0.77, 2.38]	0.03	0.00	< .001
Life Meaning	283,403	6.01	[5.84, 6.18]	0.13	**0.02**	< .001*	262,165	5.63	[5.46, 5.81]	0.12	**0.02**	< .001*	21,238	1.76	[1.13, 2.38]	0.04	0.00	< .001*
TAPAS																		
Achievement	132,432	5.24	[4.84, 5.65]	0.07	0.00	< .001*	131,561	5.16	[4.75, 5.57]	0.07	0.00	< .001*	871	3.28	[−0.43, 6.98]	0.06	0.00	.083
Adjustment	96,486	−0.08	[−0.59, 0.43]	0.00	0.00	.769	95,634	0.05	[−0.46, 0.56]	0.00	0.00	.855	852	−3.43	[−6.82, −0.04]	−0.07	0.00	.048
Adaptation	132,432	5.72	[5.49, 5.90]	0.15	**0.02**	< .001*	131,561	5.61	[5.38, 5.78]	0.15	**0.02**	< .001*	871	4.52	[2.45, 6.32]	0.15	**0.02**	< .001*
Adventure	23,018	5.39	[4.58, 6.20]	0.09	**0.01**	< .001*	22,887	5.28	[4.46, 6.09]	0.08	**0.01**	< .001*	131	5.01	[−0.90, 10.92]	0.15	**0.02**	.096
Attention Seeking	108,539	4.06	[3.66, 4.46]	0.06	0.00	< .001*	107,680	3.93	[3.54, 4.33]	0.06	0.00	< .001*	859	4.14	[1.04, 7.25]	0.09	**0.01**	.009
Commitment to Serve	47,135	−0.74	[−1.37, −0.12]	−0.01	0.00	.020	46,990	−0.74	[−1.36, −0.11]	−0.01	0.00	.021	145	−0.81	[−7.49, 5.87]	−0.02	0.00	.811
Cooperation	85,571	−1.02	[−1.51, −0.49]	−0.01	0.00	< .001	84,842	−0.88	[−1.37, −0.34]	−0.01	0.00	.001	729	−0.52	[−4.94, 3.91]	−0.01	0.00	.819
Courage	34,808	2.00	[1.30, 2.70]	0.03	0.00	< .001*	34,673	1.99	[1.29, 2.69]	0.03	0.00	< .001*	135	1.77	[−9.67, 13.20]	0.03	0.00	.760
Dominance	132,432	3.88	[3.52, 4.24]	0.06	0.00	< .001*	131,561	3.69	[3.33, 4.05]	0.05	0.00	< .001*	871	3.87	[0.44, 7.30]	0.07	**0.01**	.027
Even Tempered	132,432	−1.24	[−1.63, −0.86]	−0.02	0.00	< .001*	131,561	−1.14	[−1.53, −0.76]	−0.02	0.00	< .001*	871	−2.18	[−6.03, 1.68]	−0.04	0.00	.268
Intellectual Efficiency	132,432	0.11	[−0.26, 0.48]	0.00	0.00	.550	131,561	0.09	[−0.28, 0.46]	0.00	0.00	.643	871	−2.11	[−5.40, 1.18]	−0.04	0.00	.208
Non-Delinquency	108,315	−2.29	[−2.73, −1.84]	−0.03	0.00	< .001*	107,458	−2.24	[−2.68, −1.79]	−0.03	0.00	< .001*	857	−3.96	[−8.02, −0.64]	−0.08	**0.01**	.021
Optimism	132,432	2.34	[1.90, 2.79]	0.03	0.00	< .001*	131,561	2.32	[1.87, 2.77]	0.03	0.00	< .001*	871	−0.58	[−4.38, 3.27]	−0.01	0.00	.776
Order	109,453	2.04	[1.64, 2.43]	0.03	0.00	< .001*	108,712	2.14	[1.74, 2.54]	0.03	0.00	< .001*	741	−0.11	[−3.10, 2.89]	0.00	0.00	.944
Physical Condition	132,432	13.70	[13.40, 14.08]	0.21	**0.05**	< .001*	131,561	13.50	[13.19, 13.87]	0.21	**0.04**	< .001*	871	11.00	[7.94, 14.03]	0.23	**0.06**	< .001*
Responsibility	35,043	0.08	[−0.73, 0.84]	0.00	0.00	.893	34,906	−0.00	[−0.81, 0.75]	0.00	0.00	.942	137	−4.79	[−16.95, 7.36]	−0.07	0.00	.437
Sociability	109,414	−0.35	[−0.72, 0.03]	−0.01	0.00	.069	108,674	−0.42	[−0.79, −0.04]	−0.01	0.00	.030	740	1.65	[−1.80, 5.10]	0.03	0.00	.347
Self-Control	73,468	0.26	[−0.27, 0.79]	0.00	0.00	.337	72,747	0.21	[−0.32, 0.74]	0.00	0.00	.437	721	−0.02	[−3.87, 3.83]	0.00	0.00	.994
Situational Awareness	34,847	−0.27	[−1.04, 0.50]	0.00	0.00	.493	34,711	−0.19	[−0.96, 0.58]	0.00	0.00	.632	136	−5.00	[−13.20, 3.20]	−0.10	**0.01**	.230
Selflessness	109,453	−0.65	[−1.13, −0.18]	−0.01	0.00	.007	108,712	−0.58	[−1.06, −0.11	−0.01	0.00	.017	741	1.72	[−2.20, 5.64]	0.03	0.00	.389
Team Orientation	35,032	1.76	[1.00, 2.52]	0.02	0.00	< .001*	34,895	1.82	[1.06, 2.58]	0.03	0.00	< .001*	137	−3.16	[−16.24, 9.93]	−0.04	0.00	.634
Tolerance	109,414	−0.94	[−1.34, −0.56]	−0.01	0.00	< .001*	108,674	−0.81	[−1.20, −0.42]	−0.01	0.00	< .001*	740	−1.30	[−4.76, 2.16]	−0.03	0.00	.461

Note. APFT = Army Physical Fitness Test, GAT = Global Assessment Tool, TAPAS = Tailor Adaptive Personality Assessment System, n = sample size, b = unstandardized coefficient, CI = confidence interval, β = standardized coefficient (equivalent to r), p = p-value, *$p < .05$ after applying family-wise Bonferroni correction ($\alpha/882 = 0.000057$); r^2 values in bold represent effect sizes greater than trivial according to Cohen (1988).

Exhibit 22.

*Plot of Simple Relationship Strength between Pr***edictor variables and APFT Score**



Note. Shaded area represents trivial effect size according to Cohen (1988).

Award Count

The descriptive statistics showed that skewness was low for the Award Count variable and a Breusch-Pagan test of heteroscedasticity showed minimal violations, both suggesting that the variable approximated a normal distribution. Thus, although the Award Count outcome variable is a count variable it was treated as having a normal distribution for analyses. To examine the simple relationships of Award Count, a series of ANOVAs were conducted for categorical predictors and simple linear regressions were conducted for continuous predictors. A summary of standardized effect sizes for each predictor is presented in Exhibit 24.

Categorical Predictors. For a summary of results, see Exhibit 22. Only Rank Group had a small effect. For Rank Group, officers (M = 3.46, SD = 3.08) had more awards than enlisted Soldiers (M = 2.01, SD = 2.68), h^2 = 0.02, 90% CI [0.02, 0.02]. When splitting on Rank Group, enlisted Soldiers showed similar patterns on the other categorical variables as the overall sample. For officers, there was a small effect of MOS Type such that Combat Arms Soldiers (M = 3.70, SD = 3.13) had the highest number of awards while Operations Soldiers had the lowest number (M = 0.92, SD = 1.55).

Continuous Predictors. For a summary of results, see Exhibit 23. Several continuous predictors showed small effects with Award Count. Positive relationships included AFQT Score (+), Adaptability (+), Active Coping (+), Character (+), Optimism (+; GAT), Positive Affect (+), Organizational Trust (+), Work Engagement (+), Life Meaning (+), Dominance (+), and Physical Condition (+). Negative relationships included Catastrophizing (−), Depression (−), Negative Affect (−), and Loneliness (−). Overall, Life Meaning had the strongest relationship with an r^2 value of 0.03. Similar patterns were found when looking at only enlisted Soldiers. For officers, similar effect sizes were observed for the TAPAS measures (with Physical Condition not reaching the small effect criteria) but only Work Engagement and Life Meaning from the GAT reached the level of a small effect. However, additional patterns were found for officers including a small effect of Adventure (+) and Age at Accession (−).

Exhibit 23.

Categorical Predictors with Award Count by Sample Type

Predictors	*n*	*F*	*h*[2]	90 % CI	*p*	Group *M*s (*SD*s)
Enlisted + Officer						
Rank Group	533,317	8,896.26	**0.02**	[0.02, 0.02]	< .001*	*Enlisted*: 2.01 (2.68) *Officer*: 3.46 (3.08)
MOS Type	533,226	186.95	0.00	[0.00, 0.00]	< .001*	*Combat Arms*: 2.14 (2.81) *Combat Support*: 1.97 (2.61) *Combat Service Support*: 2.02 (2.63) *Special Service*: 2.10 (2.63) *Operations*: 1.46 (2.59) *Unknown*: 2.47 (2.82)
Soldier Sex	533,317	507.60	0.00	[0.00, 0.00]	< .001*	*Male*: 2.14 (2.76) *Female*: 1.91 (2.58)
Soldier Race	533,317	385.20	0.00	[0.00, 0.00]	< .001*	*White*: 2.11 (2.78) *Black*: 1.93 (2.46) *Asian*: 2.64 (2.95) *American Indian/Alaskan Native*: 1.78 (2.51) *Native Hawaiian/Pacific Islander*: 2.09 (2.63) *Mixed Race/Other*: 2.78 (3.10)
Enlisted						
Rank Group	—	—	—	—	—	—
MOS Type	499,801	249.43	0.00	[0.00, 0.00]	< .001*	*Combat Arms*: 2.04 (2.76) *Combat Support*: 1.87 (2.54) *Combat Service Support*: 1.96 (2.59) *Special Service*: 1.91 (2.53) *Operations*: 1.49 (2.64) *Unknown*: 2.46 (2.81)
Soldier Sex	499,807	726.84	0.00	[0.00, 0.00]	< .001*	*Male*: 2.06 (2.71) *Female*: 1.77 (2.49)
Soldier Race	499,807	301.11	0.00	[0.00, 0.00]	< .001*	*White*: 2.01 (2.73) *Black*: 1.91 (2.45) *Asian*: 2.56 (2.92) *American Indian/Alaskan Native*: 1.72 (2.47) *Native Hawaiian/Pacific Islander*: 3.06 (2.63) *Mixed Race/Other*: 1.53 (2.72)
Officer						
Rank Group	—	—	—	—	—	—
MOS Type	33,419	63.25	**0.01**	[0.01, 0.01]	< .001*	*Combat Arms*: 3.70 (3.13) *Combat Support*: 3.58 (3.18) *Combat Service Support*: 3.48 (3.11) *Special Service*: 3.02 (2.90) *Operations*: 0.92 (1.55) *Unknown*: 3.56 (3.12)
Soldier Sex	33,508	37.33	0.00	[0.00, 0.00]	< .001*	*Male*: 3.51 (3.10) *Female*: 3.27 (2.99)
Soldier Race	33,510	14.63	0.00	[0.00, 0.00]	< .001*	*White*: 3.50 (3.11) *Black*: 2.95 (2.71) *Asian*: 3.38 (3.07) *American Indian/Alaskan Native*: 3.40 (2.87) *Native Hawaiian/Pacific Islander*: 3.09 (2.49) *Mixed Race/Other*: 3.54 (3.08)

Note. *n* = sample size, *F* = *F*-test coefficient, h^2= eta squared, CI = confidence interval of effect size, *p* = p-value, *$p < .05$ after applying family-wise Bonferroni correction (α/882 = 0.000057); effect size values in bold represent effect sizes greater than trivial according to Cohen (1988).

Exhibit 24.

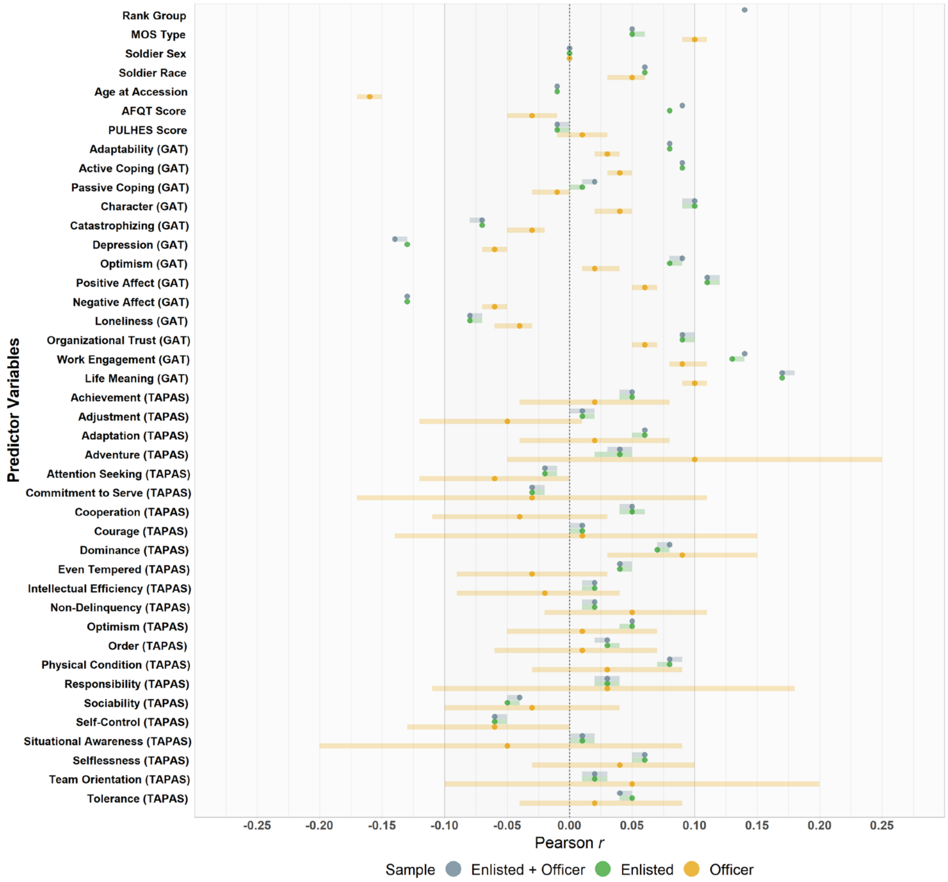
Continuous Predictors with Award Count by Sample Type

Predictors	Enlisted + Officer						Enlisted						Officer					
	n	*b*	95% CI	*β*	r^2	*p*	*n*	*b*	95% CI	*β*	r^2	*p*	*n*	*b*	95% CI	*β*	r^2	*p*
Accession																		
Age at Accession	533,317	−0.01	[−0.01, −0.01]	−0.01	0.00	< .001*	499,807	−0.01	[−0.01, −0.01]	−0.01	0.00	< .001*	33,510	−0.10	[−0.11, −0.09]	−0.16	**0.02**	< .001*
AFQT Score	507,927	0.01	[0.01, 0.01]	0.09	**0.01**	< .001*	498,515	0.01	[0.01, 0.01]	0.08	**0.01**	< .001*	9,412	−0.01	[−0.01, 0.00]	−0.03	0.00	.004
PULHES Score	497,529	−0.10	[−0.16, −0.04]	−0.01	0.00	< .001	489,371	−0.13	[−0.19, −0.07]	−0.01	0.00	< .001*	8,158	0.13	[−0.22, 0.48]	0.01	0.00	.477
GAT																		
Adaptability	349,464	0.29	[0.28, 0.30]	0.08	**0.01**	< .001*	324,727	0.29	[0.27, 0.30]	0.08	**0.01**	< .001*	24,737	0.13	[0.07, 0.18]	0.03	0.00	< .001*
Active Coping	349,455	0.34	[0.33, 0.35]	0.09	**0.01**	< .001*	324,718	0.33	[0.32, 0.34]	0.09	**0.01**	< .001*	24,737	0.20	[0.14, 0.26]	0.04	0.00	< .001*
Passive Coping	349,455	0.05	[0.04, 0.06]	0.02	0.00	< .001*	324,718	0.02	[0.01, 0.03]	0.01	0.00	.002	24,737	−0.05	[−0.09, 0.00]	−0.01	0.00	.030
Character	349,454	0.18	[0.17, 0.19]	0.10	**0.01**	< .001*	324,720	0.18	[0.17, 0.18]	0.10	**0.01**	< .001*	24,734	0.09	[0.06, 0.12]	0.04	0.00	< .001*
Catastrophizing	349,463	−0.25	[−0.27, −0.24]	−0.07	**0.01**	< .001*	324,727	−0.23	[−0.24, −0.22]	−0.07	0.00	< .001*	24,736	−0.16	[−0.21, −0.10]	−0.03	0.00	< .001*
Depression	349,459	−0.48	[−0.49, −0.47]	−0.14	**0.02**	< .001*	324,722	−0.45	[−0.46, −0.44]	−0.13	**0.02**	< .001*	24,737	−0.35	[−0.43, −0.28]	−0.06	0.00	< .001*
Optimism	336,509	0.31	[0.29, 0.32]	0.09	**0.01**	< .001*	312,602	0.28	[0.27, 0.29]	0.08	**0.01**	< .001*	23,907	0.10	[0.05, 0.16]	0.02	0.00	< .001
Positive Affect	336,508	0.40	[0.38, 0.41]	0.11	**0.01**	< .001*	312,600	0.38	[0.37, 0.39]	0.11	**0.01**	< .001*	23,908	0.29	[0.23, 0.34]	0.06	0.00	< .001*
Negative Affect	336,508	−0.50	[−0.51, −0.48]	−0.13	**0.02**	< .001*	312,600	−0.49	[−0.50, −0.47]	−0.13	**0.02**	< .001*	23,908	−0.31	[−0.38, −0.24]	−0.06	0.00	< .001*
Loneliness	349,463	−0.25	[−0.26, −0.24]	−0.08	**0.01**	< .001*	324,727	−0.24	[−0.25, −0.22]	−0.08	**0.01**	< .001*	24,736	−0.18	[−0.23, −0.13]	−0.04	0.00	< .001*
Organizational Trust	349,463	0.32	[0.31, 0.33]	0.09	**0.01**	< .001*	324,728	0.31	[0.30, 0.32]	0.09	**0.01**	< .001*	24,735	0.28	[0.22, 0.33]	0.06	0.00	< .001*
Work Engagement	224,437	0.38	[0.37, 0.40]	0.14	**0.02**	< .001*	207,807	0.35	[0.34, 0.36]	0.13	**0.02**	< .001*	16,630	0.38	[0.32, 0.44]	0.09	**0.01**	< .001*
Life Meaning	349,455	0.55	[0.54, 0.56]	0.17	**0.03**	< .001*	324,719	0.54	[0.52, 0.55]	0.17	**0.03**	< .001*	24,736	0.39	[0.34, 0.44]	0.10	**0.01**	< .001*
TAPAS																		
Achievement	181,121	0.28	[0.25, 0.30]	0.05	0.00	< .001*	180,083	0.27	[0.24, 0.30]	0.05	0.00	< .001*	1,038	0.11	[−0.25, 0.47]	0.02	0.00	.548
Adjustment	124,226	0.06	[0.03, 0.10]	0.01	0.00	< .001	123,216	0.07	[0.04, 0.10]	0.01	0.00	< .001*	1,010	−0.30	[−0.65, 0.04]	−0.05	0.00	.086
Adaptation	181,121	0.17	[0.16, 0.19]	0.06	0.00	< .001*	180,083	0.17	[0.16, 0.18]	0.06	0.00	< .001*	1,038	0.06	[−0.12, 0.26]	0.02	0.00	.482
Adventure	29,855	0.19	[0.13, 0.25]	0.04	0.00	< .001*	29,681	0.18	[0.12, 0.24]	0.04	0.00	< .001*	174	0.50	[−0.23, 1.22]	0.10	**0.01**	.176
Attention Seeking	143,243	−0.09	[−0.11, −0.06]	−0.02	0.00	< .001*	142,223	−0.09	[−0.12, −0.07]	−0.02	0.00	< .001*	1,020	−0.29	[−0.60, 0.02]	−0.06	0.00	.070
Commitment to Serve	67,915	−0.15	[−0.19, −0.11]	−0.03	0.00	< .001*	67,722	−0.15	[−0.19, −0.11]	−0.03	0.00	< .001*	193	−0.15	[−0.81, 0.52]	−0.03	0.00	.670
Cooperation	113,502	0.27	[0.24, 0.31]	0.05	0.00	< .001*	112,653	0.28	[0.25, 0.32]	0.05	0.00	< .001*	849	−0.28	[−0.76, 0.21]	−0.04	0.00	.262
Courage	48,602	0.31	[−0.02, 0.08]	0.01	0.00	.191	48,423	0.03	[−0.02, 0.08]	0.01	0.00	.197	179	0.02	[−0.60, 0.64]	0.01	0.00	.945
Dominance	181,121	0.40	[0.37, 0.42]	0.08	**0.01**	< .001*	180,083	0.39	[0.36, 0.41]	0.07	**0.01**	< .001*	1,038	0.52	[0.18, 0.85]	0.09	**0.01**	.002
Even Tempered	181,121	0.24	[0.21, 0.26]	0.04	0.00	< .001*	180,083	0.24	[0.22, 0.27]	0.04	0.00	< .001*	1,038	−0.17	[−0.55, 0.21]	−0.03	0.00	.378
Intellectual Efficiency	181,121	0.10	[0.07, 0.12]	0.02	0.00	< .001*	180,083	0.10	[0.07, 0.12]	0.02	0.00	< .001*	1,038	−0.13	[−0.46, 0.19]	−0.02	0.00	.428
Non-Delinquency	143,061	0.09	[0.06, 0.12]	0.02	0.00	< .001*	142,042	0.09	[0.06, 0.12]	0.02	0.00	< .001*	1,019	0.29	[−0.09, 0.64]	0.05	0.00	.142
Optimism	181,121	0.32	[0.29, 0.35]	0.05	0.00	< .001*	180,083	0.32	[0.29, 0.35]	0.05	0.00	< .001*	1,038	0.04	[−0.34, 0.42]	0.01	0.00	.844
Order	151,354	0.15	[0.12, 0.18]	0.03	0.00	< .001*	150,486	0.16	[0.13, 0.18]	0.03	0.00	< .001*	868	0.03	[−0.30, 0.36]	0.01	0.00	.862
Physical Condition	181,121	0.40	[0.37, 0.42]	0.08	**0.01**	< .001*	180,083	0.39	[0.36, 0.41]	0.08	**0.01**	< .001*	1,038	0.16	[−0.15, 0.47]	0.03	0.00	.311
Responsibility	48,810	0.19	[0.13, 0.24]	0.03	0.00	< .001*	48,631	0.19	[0.13, 0.24]	0.03	0.00	< .001*	179	0.17	[−0.54, 0.87]	0.03	0.00	.645
Sociability	151,266	−0.21	[−0.24, −0.19]	−0.04	0.00	< .001*	150,402	−0.22	[−0.24, −0.19]	−0.05	0.00	< .001*	864	−0.15	[−0.48, 0.19]	−0.03	0.00	.390
Self-Control	94,371	−0.30	[−0.33, −0.27]	−0.06	0.00	< .001*	93,535	−0.30	[−0.34, −0.27]	−0.06	0.00	< .001*	836	−0.36	[−0.74, 0.03]	−0.06	0.00	.067
Situational Awareness	48,690	0.06	[0.00, 0.11]	0.01	0.00	.037	48,507	0.06	[0.01, 0.11]	0.01	0.00	.023	183	−0.31	[−1.16, 0.54]	−0.05	0.00	.470
Selflessness	151,354	0.35	[0.32, 0.38]	0.06	0.00	< .001*	150,486	0.35	[0.32, 0.38]	0.06	0.00	< .001*	868	0.25	[−0.18, 0.68]	0.04	0.00	.255
Team Orientation	48,784	0.13	[0.08, 0.18]	0.02	0.00	< .001*	48,604	0.13	[0.08, 0.18]	0.02	0.00	< .001*	180	0.26	[−0.49, 1.01]	0.05	0.00	.502
Tolerance	151,266	0.22	[0.19, 0.24]	0.04	0.00	< .001*	150,402	0.23	[0.20, 0.25]	0.05	0.00	< .001*	864	0.12	[−0.22, 0.45]	0.02	0.00	.489

Note. GAT = Global Assessment Tool, TAPAS = Tailor Adaptive Personality Assessment System, *n* = sample size, *b* = unstandardized coefficient, CI = confidence interval, *β* = standardized coefficient (equivalent to *r*), *p* = p-value, **p* < .05 after applying family-wise Bonferroni correction (α/882 = 0.000057); r^2 values in bold represent effect sizes greater than trivial according to Cohen (1988).

Exhibit 25.

Plot of Simple Relationship Strength between Predictor Variables and Award Count



Note. Shaded area represents trivial effect size according to Cohen (1988).

Bad Paper Count

The descriptive statistics showed that skewness was moderate for the Award Count variable and a Breusch-Pagan test of heteroscedasticity showed moderate violations, both suggesting that the variable did not closely approximate a normal distribution. Thus, the Bad Paper Count variable was treated as count data and modeled using a Poisson distribution. To examine the simple relationships of Bad Paper Count, a series of generalized linear models with a log linkage were used for both categorical predictors and continuous predictors. A summary of standardized effect sizes for each predictor is presented in Exhibit 27.

Categorical Predictors. For a summary of results, see Exhibit 25. Only Soldier Sex had a small effect. For Soldier Sex, male Soldiers (M = 0.04, SD = 0.23) were more likely to receive a bad paper than female Soldiers (M = 0.02, SD = 0.16), OR = 2.02, 95% CI [1.92, 2.12]. Similar patterns were found when looking at only enlisted Soldiers but not for officers.

Continuous Predictors. For a summary of results, see Exhibit 26. No relationships emerged at or above the small effect criteria when considering enlisted Soldiers and officers together. This pattern was the same when only looking at enlisted Soldiers. For officers, however, Intellectual Efficiency (+), Responsibility (−), and Team Orientation (−) met or passed the small effect size criteria.

Exhibit 26.

Categorical Predictors with Bad Paper Count by Sample Type

Predictors		*n*	*z*	*OR*	95 % CI	*p*	Group *M*s (*SD*s)
Enlisted + Officer							
	Rank Group	533,317	−0.71	0.98	[0.93, 1.04]	.478	*Enlisted*: 0.04 (0.22) *Officer*: 0.04 (0.22)
	MOS Type	533,226	*F* = 21.06	1.05	[1.05, 1.05]	< .001*	*Combat Arms*: 0.04 (0.22) *Combat Support*: 0.04 (0.22) *Combat Service Support*: 0.04 (0.23) *Special Service*: 0.04 (0.22) *Operations*: 0.07 (0.32) *Unknown*: 0.03 (0.19)
	Soldier Sex	533,317	28.61	**2.02**	[1.92, 2.12]	< .001*	*Male*: 0.04 (0.23) *Female*: 0.02 (0.16)
	Soldier Race	533,317	*F* = 157.93	1.28	[1.28, 1.28]	< .001*	*White*: 0.04 (0.21) *Black*: 0.06 (0.26) *Asian*: 0.03 (0.17) *American Indian/Alaskan Native*: 0.06 (0.27) *Native Hawaiian/Pacific Islander*: 0.03 (0.17) *Mixed Race/Other*: 0.07 (0.28)
Enlisted							
	Rank Group	—	—	—	—	—	—
	MOS Type	499,807	*F* = 19.19	0.00	[0.00, 0.00]	< .001*	*Combat Arms*: 0.04 (0.22) *Combat Support*: 0.04 (0.22) *Combat Service Support*: 0.04 (0.22) *Special Service*: 0.04 (0.22) *Operations*: 0.07 (0.30) *Unknown*: 0.03 (0.19)
	Soldier Sex	499,807	29.50	**2.19**	[2.08, 2.31]	< .001*	*Male*: 0.04 (0.23) *Female*: 0.02 (0.15)
	Soldier Race	499,807	*F* = 141.02	1.15	[1.15, 1.15]	< .001*	*White*: 0.04 (0.21) *Black*: 0.05 (0.26) *Asian*: 0.03 (0.17) *American Indian/Alaskan Native*: 0.06 (0.27) *Native Hawaiian/Pacific Islander*: 0.03 (0.17) *Mixed Race/Other*: 0.08 (0.31)
Officer							
	Rank Group	—	—	—	—	—	—
	MOS Type	33,419	*F* = 12.70	1.15	[1.12, 1.22]	< .001*	*Combat Arms*: 0.04 (0.21) *Combat Support*: 0.04 (0.23) *Combat Service Support*: 0.06 (0.27) *Special Service*: 0.03 (0.20) *Operations*: 0.16 (0.50) *Unknown*: 0.04 (0.19)
	Soldier Sex	33,510	0.74	1.05	[0.92, 1.20]	.461	*Male*: 0.04 (0.22) *Female*: 0.04 (0.21)
	Soldier Race	33,510	*F* = 31.18	1.29	[1.22, 1.33]	< .001*	*White*: 0.03 (0.21) *Black*: 0.09 (0.33) *Asian*: 0.04 (0.21) *American Indian/Alaskan Native*: 0.04 (0.20) *Native Hawaiian/Pacific Islander*: 0.06 (0.25) *Mixed Race/Other*: 0.06 (0.27)

Note. *n* = sample size, *F* = *F*-test coefficient, *OR* = odds ratio, CI = confidence interval of effect size, *p* = p-value, **p* < .05 after applying family-wise Bonferroni correction (α/882 = 0.000057); effect size values in bold represent effect sizes greater than trivial according to Chen et al. (2010).

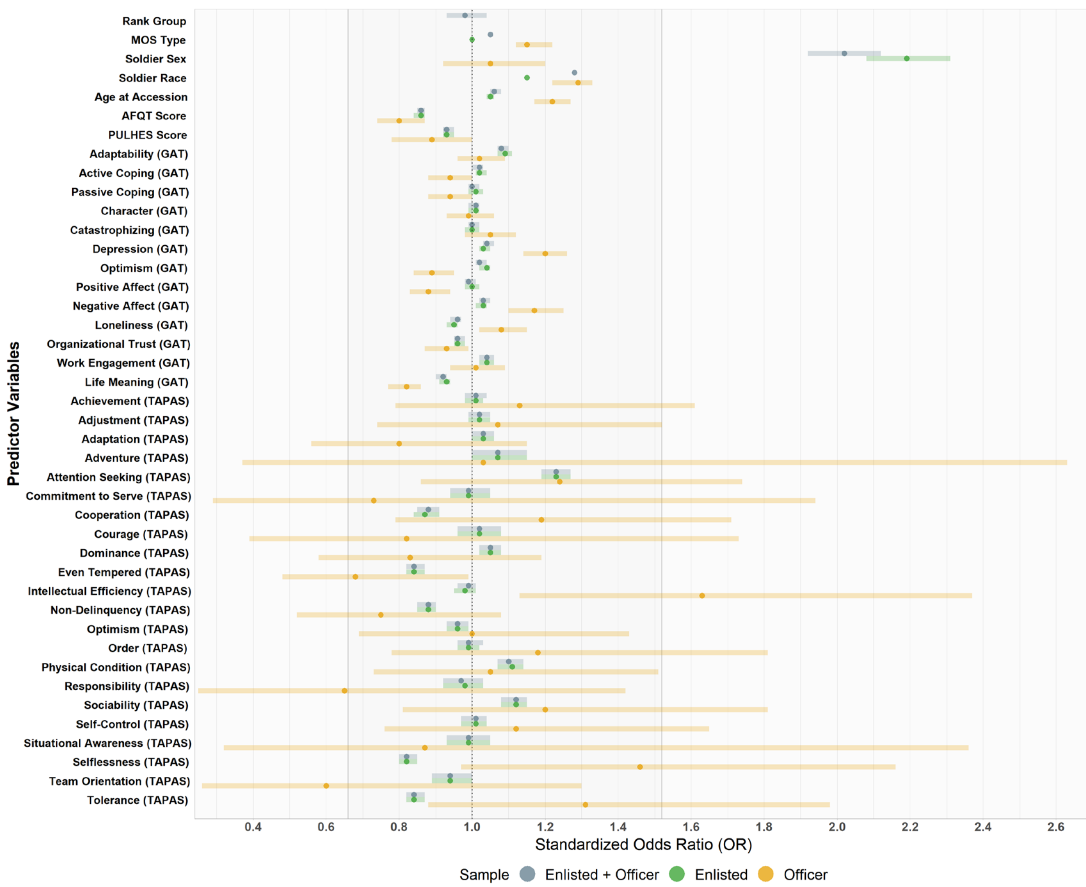
Exhibit 27.

Continuous Predictors with Bad Paper Count by Sample Type

	Enlisted + Officer						*Enlisted*						*Officer*					
Predictors	***n***	***b***	**95% CI**	***β***	***OR***	***p***	***n***	***b***	**95% CI**	***β***	***OR***	***p***	***n***	***b***	**95% CI**	***β***	***OR***	***p***
Accession																		
Age at Accession	533,317	0.01	[0.01, 0.02]	0.06	1.06	< .001*	498,515	0.01	[0.01, 0.01]	0.05	1.05	< .001*	33,510	0.04	[0.03, 0.05]	0.20	1.22	< .001*
AFQT Score	507,927	−0.01	[−0.01, −0.01]	−0.17	0.86	< .001*	489,371	−0.01	[−0.01, −0.01]	−0.16	0.86	< .001*	9,412	0.00	[−0.02, −0.01]	−0.22	0.80	< .001*
PULHES Score	497,529	−0.54	[−0.65, −0.42]	−0.07	0.93	< .001*	489,371	−0.54	[−0.66, −0.42]	−0.07	0.93	< .001*	8,158	−0.58	[−1.28, 0.01]	−0.11	0.89	.079
GAT																		
Adaptability	349,464	0.11	[0.08, 0.13]	0.08	1.08	< .001*	324,727	0.11	[0.09, 0.13]	0.09	1.09	< .001*	24,737	0.03	[−0.06, 0.13]	0.02	1.02	.494
Active Coping	349,455	0.02	[0.00, 0.04]	0.02	1.02	.050	324,718	0.03	[0.01, 0.05]	0.02	1.02	.011	24,737	−0.11	[−0.21, 0.00]	−0.07	0.94	.042
Passive Coping	349,455	0.00	[−0.01, 0.02]	0.00	1.00	.735	324,718	0.01	[−0.01, 0.03]	0.01	1.01	.243	24,737	−0.07	[−0.15, 0.00]	−0.06	0.94	.053
Character	349,454	0.00	[−0.01, 0.01]	0.01	1.01	.538	324,720	0.00	[−0.01, 0.02]	0.01	1.01	.450	24,734	−0.01	[−0.06, 0.05]	−0.01	0.99	.843
Catastrophizing	349,463	0.00	[−0.02, 0.02]	0.00	1.00	.676	324,727	−0.00	[−0.02, 0.02]	−0.00	1.00	.916	24,736	0.07	[−0.02, 0.16]	0.05	1.05	.136
Depression	349,459	0.05	[0.03, 0.07]	0.04	1.04	< .001*	324,722	0.04	[0.02, 0.06]	0.03	1.03	< .001	24,737	0.35	[0.26, 0.44]	0.18	1.20	< .001*
Optimism	336,509	0.03	[0.01, 0.05]	0.02	1.02	.004	312,602	0.04	[0.02, 0.07]	0.04	1.04	< .001	23,907	−0.17	[−0.26, −0.08]	−0.12	0.89	< .001
Positive Affect	336,508	−0.01	[−0.03, 0.01]	−0.01	0.99	.416	312,600	0.00	[−0.02, 0.02]	0.00	1.00	.869	23,908	−0.19	[−0.28, −0.10]	−0.13	0.88	< .001
Negative Affect	336,508	0.05	[0.02, 0.07]	0.03	1.03	< .001	312,600	0.03	[0.01, 0.06]	0.02	1.03	.004	23,908	0.28	[0.17, 0.39]	0.16	1.17	< .001*
Loneliness	349,463	−0.05	[−0.07, −0.03]	−0.05	0.96	< .001*	324,727	−0.06	[−0.08, −0.04]	−0.05	0.95	< .001*	24,736	0.11	[0.02, 0.19]	0.08	1.08	.013
Organizational Trust	349,463	−0.05	[−0.07, −0.03]	−0.04	0.96	< .001*	324,728	−0.04	[−0.06, −0.02]	−0.04	0.96	< .001*	24,735	−0.11	[−0.20, −0.01]	−0.07	0.93	.023
Work Engagement	224,437	0.04	[0.02, 0.06]	0.04	1.04	< .001	207,807	0.04	[0.02, 0.06]	0.04	1.04	< .001*	16,630	0.02	[−0.08, 0.11]	0.01	1.01	.731
Life Meaning	349,455	−0.10	[−0.12, −0.08]	−0.09	0.92	< .001*	324,719	−0.09	[−0.11, −0.07]	−0.08	0.93	< .001*	24,736	−0.27	[−0.34, −0.19]	−0.20	0.82	< .001*
TAPAS																		
Achievement	181,121	0.01	[−0.05, 0.07]	0.01	1.01	.675	180,083	0.01	[−0.05, 0.07]	0.01	1.01	.716	1,038	0.25	[−0.49, 0.98]	0.12	1.13	.502
Adjustment	124,226	0.04	[−0.02, 0.11]	0.02	1.02	.217	123,216	0.04	[−0.03, 0.11]	0.02	1.02	.230	1,010	0.13	[−0.58, 0.80]	0.07	1.07	.718
Adaptation	181,121	0.03	[0.00, 0.06]	0.03	1.03	.068	180,083	0.03	[0.00, 0.06]	0.03	1.03	.056	1,038	−0.24	[−0.63, 0.15]	−0.23	0.80	.220
Adventure	29,855	0.11	[0.00, 0.23]	0.07	1.07	.054	29,681	0.11	[0.00, 0.23]	0.07	1.07	.053	174	0.05	[−1.73, 1.68]	0.03	1.03	.951
Attention Seeking	143,243	0.37	[0.31, 0.42]	0.21	1.23	< .001*	142,223	0.37	[0.31, 0.42]	0.21	1.23	< .001*	1,020	0.37	[−0.25, 0.96]	0.21	1.24	.234
Commitment to Serve	67,915	−0.02	[−0.12, 0.09]	−0.01	0.99	.759	67,722	−0.01	[−0.11, 0.09]	−0.01	0.99	.790	193	−0.54	[−2.09, 1.12]	−0.32	0.73	.502
Cooperation	113,502	−0.27	[−0.34, −0.20]	−0.13	0.88	< .001*	112,653	−0.27	[−0.34, −0.20]	−0.13	0.87	< .001*	849	0.42	[−0.56, 1.28]	0.18	1.19	.375
Courage	48,602	0.04	[−0.07, 0.14]	0.02	1.02	.525	48,423	0.04	[−0.07, 0.15]	0.02	1.02	.497	179	−0.36	[−1.70, 0.97]	−0.20	0.82	.597
Dominance	181,121	0.09	[0.04, 0.14]	0.05	1.05	< .001	180,083	0.09	[0.04, 0.14]	0.05	1.05	< .001	1,038	−0.35	[−1.04, 0.33]	−0.19	0.83	.315
Even Tempered	181,121	−0.33	[−0.38, −0.27]	−0.17	0.84	< .001*	180,083	−0.33	[−0.38, −0.27]	−0.17	0.84	< .001*	1,038	−0.81	[−1.59, −0.03]	−0.38	0.68	.042
Intellectual Efficiency	181,121	−0.03	[−0.08, 0.03]	−0.01	0.99	.307	180,083	−0.03	[−0.09, 0.02]	−0.02	0.98	.218	1,038	0.90	[0.22, 1.58]	0.49	**1.63**	.009
Non-Delinquency	143,061	−0.26	[−0.32, −0.20]	−0.13	0.88	< .001*	142,042	−0.26	[−0.32, −0.20]	−0.13	0.88	< .001*	1,019	−0.59	[−1.35, 0.17]	−0.29	0.75	.128
Optimism	181,121	−0.10	[−0.16, −0.03]	−0.04	0.96	.003	180,083	−0.10	[−0.16, −0.03]	−0.04	0.96	.003	1,038	0.00	[−0.79, 0.77]	0.00	1.00	.997
Order	151,354	−0.01	[−0.07, 0.05]	−0.01	0.99	.721	150,486	−0.01	[−0.07, 0.04]	−0.01	0.99	.673	868	0.28	[−0.42, 0.99]	0.16	1.18	.445
Physical Condition	181,121	0.17	[0.12, 0.22]	0.10	1.10	< .001*	180,083	0.17	[0.12, 0.22]	0.10	1.11	< .001*	1,038	0.08	[−0.55, 0.72]	0.05	1.05	.805
Responsibility	48,810	−0.05	[−0.17, 0.07]	−0.03	0.97	.392	48,631	−0.05	[−0.17, 0.07]	−0.02	0.98	.422	179	−0.86	[−2.74, 0.70]	−0.43	**0.65**	.324
Sociability	151,266	0.19	[0.14, 0.24]	0.11	1.12	< .001*	150,402	0.19	[0.13, 0.24]	0.11	1.12	< .001*	864	0.32	[−0.36, 1.01]	0.18	1.20	.365
Self-Control	94,371	0.01	[−0.05, 0.08]	0.01	1.01	.693	93,535	0.01	[−0.06, 0.08]	0.01	1.01	.729	836	0.22	[−0.53, 0.96]	0.12	1.12	.558
Situational Awareness	48,690	−0.02	[−0.14, 0.10]	−0.01	0.99	.699	48,507	−0.02	[−0.14. 0.10]	−0.01	0.99	.709	183	−0.28	[−2.43, 1.82]	−0.13	0.87	.790
Selflessness	151,354	−0.42	[−0.49, −0.35]	−0.19	0.82	< .001*	150,486	−0.43	[−0.50, −0.36]	−0.20	0.82	< .001*	868	0.81	[−0.06, 1.65]	0.38	1.46	.062
Team Orientation	48,784	−0.12	[−0.24, −0.01]	−0.06	0.94	.039	48,604	−0.12	[−0.23, 0.00]	−0.06	0.94	.050	180	−1.08	[−2.80, 0.55]	−0.51	**0.60**	.212
Tolerance	151,266	−0.30	[−0.36, −0.25]	−0.17	0.84	< .001*	150,402	−0.31	[−0.36, −0.25]	−0.17	0.84	< .001*	864	0.46	[−0.23, 1.17]	0.27	1.31	.197

Note. GAT = Global Assessment Tool, TAPAS = Tailor Adaptive Personality Assessment System, n = sample size, b = log odds ratio coefficient, CI = confidence interval, β = standardized log odds ratio coefficient, *OR* = odds ratio, p = p-value, *$p < .05$ after applying family-wise Bonferroni correction (α/882 = 0.000057); *OR* values in bold represent effect sizes greater than trivial according to Chen et al. (2010).

Exhibit 28.
Plot of Simple Relationship Strength between Predictor Variables and Bad Paper Count



Note. Shaded area represents trivial effect size according to Cohen (1988).

Speed of Promotion

To examine the simple relationships of Speed of Promotion, a series of ANOVAs were conducted for categorical predictors and simple linear regressions were conducted for continuous predictors. A summary of standardized effect sizes for each predictor is presented in Exhibit 30.

Categorical Predictors. For a summary of results, see Exhibit 28. No relationships emerged at or above the small effect criteria when considering enlisted Soldiers and officers together. This pattern was the same when only looking at enlisted Soldiers. For officers, however, there was a small effect of MOS Type such that Operations (M = −2.34, SD = 1.78) had the fastest time to promotion to a Soldiers highest rank and Unknown had the slowest (M = 0.23, SD = 1.24).

Continuous Predictors. For a summary of results, see Exhibit 29. The only continuous predictor to show a small effect with Speed of Promotion was Age at Accession (−). For enlisted Soldiers, only Physical Condition (−) showed a small effect. For officers, a moderate effect was observed for Age at Accession (−) and small effects were observed for Positive Affect (+), Organizational Trust (+), Work Engagement (+), Life Meaning (+), Responsibility (+), AFQT Score (−), Negative Affect (−), and Team Orientation (−).

Exhibit 29.

Categorical Predictors with Speed of Promotion by Sample Type

Predictors		*n*	*F*	h^2	90 % CI	*p*	Group *Ms* (*SDs*)
Enlisted + Officer							
	Rank Group	462,287	50.27	0.00	[0.00, 0.00]	< .001*	*Enlisted*: 0.00 (1.00) *Officer*: −0.04 (0.98)
	MOS Type	462,210	310.61	0.00	[0.00, 0.00]	< .001*	*Combat Arms*: −0.04 (0.97) *Combat Support*: 0.10 (1.04) *Combat Service Support*: 0.05 (1.05) *Special Service*: 0.02 (1.00) *Operations*: −0.45 (1.27) *Unknown*: −0.08 (0.95)
	Soldier Sex	462,287	121.88	0.00	[0.00, 0.00]	< .001*	*Male*: −0.01 (0.99) *Female*: 0.04 (1.04)
	Soldier Race	462,287	216.76	0.00	[0.00, 0.00]	< .001*	*White*: −0.03 (0.98) *Black*: 0.09 (1.05) *Asian*: 0.05 (0.99) *American Indian/Alaskan Native*: 0.06 (1.05) *Native Hawaiian/Pacific Islander*: 0.03 (0.99) *Mixed Race/Other*: −0.01 (1.12)
Enlisted							
	Rank Group	—	—	—	—	—	—
	MOS Type	431,475	374.35	0.00	[0.00, 0.00]	< .001*	*Combat Arms*: −0.05 (0.96) *Combat Support*: 0.11 (1.05) *Combat Service Support*: 0.06 (1.05) *Special Service*: 0.06 (1.00) *Operations*: −0.39 (1.21) *Unknown*: −0.08 (0.95)
	Soldier Sex	431,475	175.12	0.00	[0.00, 0.00]	< .001*	*Male*: −0.01 (0.99) *Female*: 0.05 (1.05)
	Soldier Race	431,475	214.73	0.00	[0.00, 0.00]	< .001*	*White*: −0.03 (0.98) *Black*: 0.09 (1.05) *Asian*: 0.06 (0.99) *American Indian/Alaskan Native*: 0.06 (1.05) *Native Hawaiian/Pacific Islander*: 0.01 (0.99) *Mixed Race/Other*: 0.00 (1.33)
Officer							
	Rank Group	—	—	—	—	—	—
	MOS Type	30,735	79.97	**0.01**	[0.01, 0.02]	< .001*	*Combat Arms*: 0.05 (1.00) *Combat Support*: −0.05 (0.93) *Combat Service Support*: −0.08 (0.94) *Special Service*: −0.19 (0.96) *Operations*: −2.34 (1.78) *Unknown*: 0.23 (1.24)
	Soldier Sex	30,812	16.08	0.00	[0.00, 0.00]	< .001*	*Male*: −0.03 (0.98) *Female*: −0.08 (0.98)
	Soldier Race	30,812	7.57	0.00	[0.00, 0.00]	< .001*	*White*: −0.04 (0.98) *Black*: 0.01 (0.97) *Asian*: −0.08 (0.99) *American Indian/Alaskan Native*: 0.08 (1.05) *Native Hawaiian/Pacific Islander*: 0.53 (0.83) *Mixed Race/Other*: −0.01 (0.99)

Note. *n* = sample size, *F* = *F*-test coefficient, h^2 = eta squared, CI = confidence interval of effect size, *p* = p-value, **p* < .05 after applying family-wise Bonferroni correction (α/882 = 0.000057); effect size values in bold represent effect sizes greater than trivial according to Cohen (1988).

Exhibit 30.

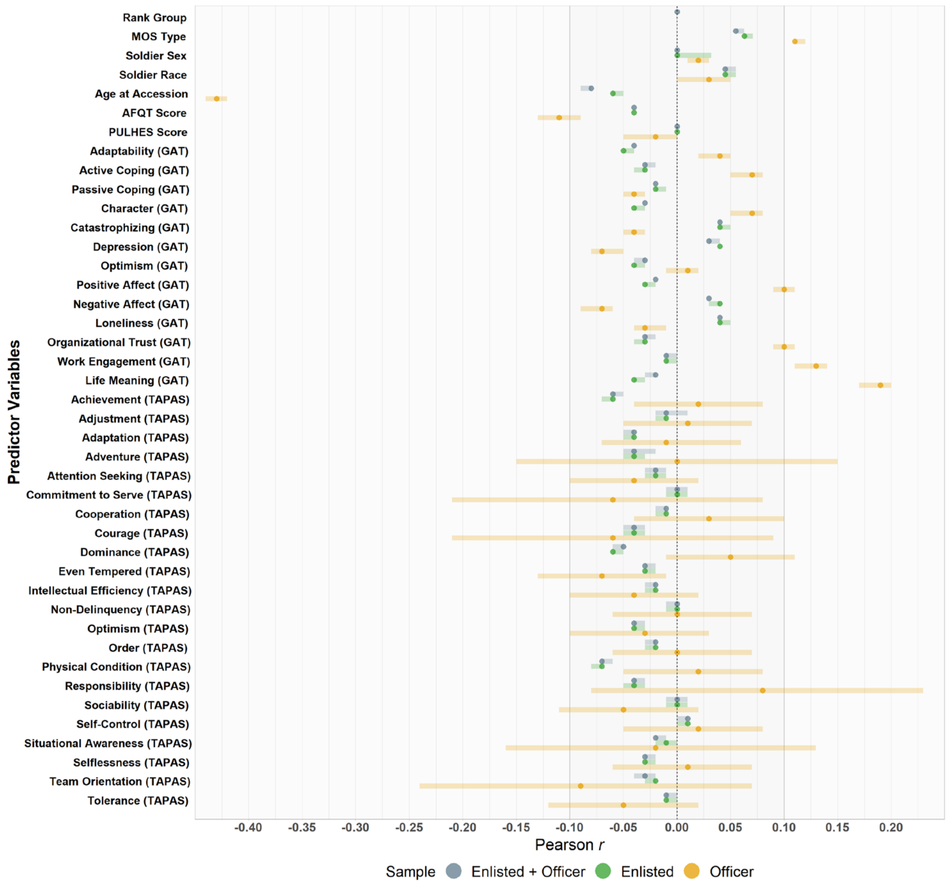
Continuous Predictors with Speed of Promotion by Sample Type

	Enlisted + Officer						Enlisted						Officer					
Predictors	***n***	***b***	**95% CI**	***β***	***r*²**	***p***	***n***	***b***	**95% CI**	***β***	***R*²**	***p***	***n***	***b***	**95% CI**	***β***	***r*²**	***p***
Accession																		
Age at Accession	462,287	−0.02	[−0.02, −0.02]	−0.08	**0.01**	< .001*	431,475	−0.01	[−0.02, −0.01]	−0.06	0.00	< .001*	30,812	−0.10	[−0.10, −0.10]	−0.43	**0.19**	< .001*
AFQT Score	439,519	−0.00	[0.00, 0.00]	−0.04	0.00	< .001*	430,431	−0.00	[0.00, 0.00]	−0.04	0.00	< .001*	9,088	−0.01	[−0.01, 0.00]	−0.11	**0.01**	< .001*
PULHES Score	431,663	−0.00	[−0.03, 0.02]	0.00	0.00	.766	423,872	0.00	[−0.02, 0.03]	0.00	0.00	< .001	7,791	−0.12	[−0.23, −0.01]	−0.02	0.00	.029
GAT																		
Adaptability	317,555	−0.05	[−0.06, −0.05]	−0.04	0.00	< .001*	294,776	−0.06	[−0.06, −0.05]	−0.05	0.00	< .001*	22,779	0.06	[0.04, 0.07]	0.04	0.00	< .001*
Active Coping	317,547	−0.04	[−0.04, −0.03]	−0.03	0.00	< .001*	294,768	−0.04	[−0.05, −0.04]	−0.03	0.00	< .001*	22,779	0.11	[0.09, 0.13]	0.07	0.00	< .001*
Passive Coping	317,547	−0.02	[−0.02, −0.02]	−0.02	0.00	< .001*	294,768	−0.02	[−0.02, −0.01]	−0.02	0.00	< .001*	22,779	−0.04	[−0.06, −0.03]	−0.04	0.00	< .001*
Character	317,549	−0.02	[−0.02, −0.02]	−0.03	0.00	< .001*	294,772	−0.02	[−0.03, −0.02]	−0.04	0.00	< .001*	22,777	0.05	[0.04, 0.07]	0.07	0.00	< .001*
Catastrophizing	317,554	0.05	[0.04, 0.05]	0.04	0.00	< .001*	294,776	0.05	[0.05, 0.06]	0.04	0.00	< .001*	22,778	−0.06	[−0.07, −0.04]	−0.04	0.00	< .001*
Depression	317,551	0.04	[0.04, 0.05]	0.03	0.00	< .001*	294,772	0.05	[0.04, 0.05]	0.04	0.00	< .001*	22,779	−0.12	[−0.15, −0.10]	−0.07	0.00	< .001*
Optimism	304,941	−0.04	[−0.05, −0.04]	−0.03	0.00	< .001*	282,948	−0.05	[−0.05, −0.04]	−0.04	0.00	< .001*	21,993	0.01	[−0.01, 0.03]	0.01	0.00	.224
Positive Affect	304,942	−0.02	[−0.03, −0.02]	−0.02	0.00	< .001*	282,948	−0.03	[−0.04, −0.03]	−0.03	0.00	< .001*	21,994	0.15	[0.13, 0.17]	0.10	**0.01**	< .001*
Negative Affect	304,942	0.04	[0.04, 0.05]	0.03	0.00	< .001*	282,948	0.05	[0.05, 0.06]	0.04	0.00	< .001*	21,994	−0.13	[−0.15, −0.10]	−0.07	**0.01**	< .001*
Loneliness	317,554	0.04	[0.04, 0.05]	0.04	0.00	< .001*	294,776	0.05	[0.04, 0.05]	0.04	0.00	< .001*	22,778	−0.03	[−0.05, −0.02]	−0.03	0.00	< .001
Organizational Trust	317,554	−0.03	[−0.04, −0.03]	−0.03	0.00	< .001*	294,777	−0.04	[−0.04, −0.04]	−0.03	0.00	< .001*	22,777	0.14	[0.12, 0.16]	0.10	**0.01**	< .001*
Work Engagement	199,870	−0.01	[−0.01, 0.00]	−0.01	0.00	.003	184,479	−0.01	[−0.01, 0.00]	−0.01	0.00	< .001	15,391	0.15	[0.13, 0.17]	0.13	**0.02**	< .001*
Life Meaning	317,547	−0.03	[−0.03, −0.02]	−0.02	0.00	< .001*	294,769	−0.04	[−0.05, −0.04]	−0.04	0.00	< .001*	22,778	0.24	[0.22, 0.25]	0.19	**0.04**	< .001*
TAPAS																		
Achievement	154,076	−0.11	[−0.12, −0.10]	−0.06	0.00	< .001*	153,065	−0.11	[−0.12, −0.10]	−0.06	0.00	< .001*	1,011	0.03	[−0.07, 0.13]	0.02	0.00	.508
Adjustment	105,237	−0.03	[−0.04, −0.02]	−0.01	0.00	< .001*	104,250	−0.03	[−0.04, −0.01]	−0.01	0.00	< .001*	987	0.01	[−0.08, 0.11]	0.01	0.00	.766
Adaptation	154,076	−0.04	[−0.04, −0.03]	−0.04	0.00	< .001*	153,065	−0.04	[−0.04, −0.04]	−0.04	0.00	< .001*	1,011	−0.03	[−0.06, 0.05]	−0.01	0.00	.863
Adventure	25,327	−0.05	[−0.07, −0.04]	−0.04	0.00	< .001*	25,155	−0.06	[−0.08, −0.04]	−0.04	0.00	< .001*	172	0.00	[−0.19, 0.19]	0.00	0.00	.974
Attention Seeking	121,613	−0.03	[−0.04, −0.02]	−0.02	0.00	< .001*	120,618	−0.03	[−0.04, −0.02]	−0.02	0.00	< .001*	995	−0.06	[−0.14, 0.03]	−0.04	0.00	.195
Commitment to Serve	58,023	0.00	[−0.01, 0.01]	0.00	0.00	.973	57,836	0.00	[−0.01, 0.01]	0.00	0.00	.980	187	−0.08	[−0.26, 0.10]	−0.06	0.00	.379
Cooperation	96,253	−0.02	[−0.03, −0.01]	−0.01	0.00	< .001	95,418	−0.02	[−0.03, −0.01]	−0.01	0.00	< .001	835	0.06	[−0.07, 0.18]	0.03	0.00	.396
Courage	41,447	−0.07	[−0.08, −0.05]	−0.04	0.00	< .001*	41,279	−0.07	[−0.08, −0.05]	−0.04	0.00	< .001*	168	−0.10	[−0.33, 0.14]	−0.06	0.00	.417
Dominance	154,076	−0.09	[−0.10, −0.08]	−0.05	0.00	< .001*	153,065	−0.09	[−0.10, −0.09]	−0.06	0.00	< .001*	1,011	0.07	[−0.02, 0.16]	0.05	0.00	.128
Even Tempered	154,076	−0.05	[−0.05, −0.04]	−0.03	0.00	< .001*	153,065	−0.04	[−0.05, −0.03]	−0.03	0.00	< .001*	1,011	−0.12	[−0.22, −0.01]	−0.07	0.00	.028
Intellectual Efficiency	154,076	−0.03	[−0.04, −0.03]	−0.02	0.00	< .001*	153,065	−0.03	[−0.04, −0.03]	−0.02	0.00	< .001*	1,011	−0.06	[−0.14, 0.03]	−0.04	0.00	.221
Non-Delinquency	121,380	−0.01	[−0.02, 0.00]	0.00	0.00	.080	120,385	−0.01	[−0.02, 0.00]	0.00	0.00	.126	996	0.01	[−0.09, 0.11]	0.00	0.00	.886
Optimism	154,076	−0.07	[−0.08, −0.06]	−0.04	0.00	< .001*	153,065	−0.07	[−0.08, −0.06]	−0.04	0.00	< .001*	1,011	−0.06	[−0.16, 0.05]	−0.03	0.00	.277
Order	128,772	−0.04	[−0.04, −0.03]	−0.02	0.00	< .001*	127,920	−0.04	[−0.04, −0.03]	−0.02	0.00	< .001*	852	0.01	[−0.08, 0.10]	0.00	0.00	.895
Physical Condition	154,076	−0.11	[−0.12, −0.10]	−0.07	0.00	< .001*	153,065	−0.11	[−0.12, −0.10]	−0.07	**0.01**	< .001*	1,011	0.02	[−0.06, 0.11]	0.02	0.00	.605
Responsibility	41,624	−0.07	[−0.08, −0.05]	−0.04	0.00	< .001*	41,458	−0.07	[−0.09, −0.05]	−0.04	0.00	< .001*	166	0.12	[−0.11, 0.35]	0.08	**0.01**	.318
Sociability	128,749	0.00	[−0.01, 0.01]	0.00	0.00	.972	127,910	0.00	[−0.01, 0.01]	0.00	0.00	.992	839	−0.06	[−0.15, 0.03]	−0.05	0.00	.190
Self-Control	79,910	0.01	[0.00, 0.02]	0.01	0.00	.061	79,095	0.01	[0.00, 0.02]	0.01	0.00	.079	815	0.02	[−0.08, 0.13]	0.02	0.00	.669
Situational Awareness	41,470	−0.03	[−0.04, −0.01]	−0.02	0.00	< .001	41,289	−0.03	[−0.04, −0.01]	−0.01	0.00	.003	181	−0.03	[−0.28, 0.22]	−0.02	0.00	.809
Selflessness	128,772	−0.05	[−0.06, −0.04]	−0.03	0.00	< .001*	127,920	−0.05	[−0.06, −0.04]	−0.03	0.00	< .001*	852	0.01	[−0.10, 0.12]	0.01	0.00	.852
Team Orientation	41,680	−0.05	[−0.06, −0.03]	−0.03	0.00	< .001*	41,513	−0.04	[−0.06, −0.03]	−0.02	0.00	< .001*	167	−0.14	[−0.38, 0.10]	−0.09	**0.01**	.268
Tolerance	128,749	−0.01	[−0.02, 0.00]	−0.01	0.00	.003	127,910	−0.01	[−0.02, 0.00]	−0.01	0.00	.016	839	−0.07	[−0.16, 0.03]	−0.05	0.00	.162

Note. GAT = Global Assessment Tool, TAPAS = Tailor Adaptive Personality Assessment System, n = sample size, b = unstandardized coefficient, CI = confidence interval, β = standardized coefficient (equivalent to r), p = p-value, *$p < .05$ after applying family-wise Bonferroni correction ($\alpha/882 = 0.000057$); r^2 values in bold represent effect sizes greater than trivial according to Cohen (1988).

Exhibit 31.

Plot of Simple Relationship Strength between Predictor Variables and Speed of Promotion



Note. Shaded area represents trivial effect size according to Cohen (1988).

First-Term Attrition

To examine the simple relationships of First-Term Attrition, a series of generalized linear models with a logit linkage were used for both categorical predictors and continuous predictors. A summary of standardized effect sizes for each predictor is presented in Exhibit 33. Overall, 35.38% (n = 175,446) attrited in their first term while 64.62% (n = 320,486) completed their first term. For enlisted Soldiers, 35.55% (n = 175,356) attrited in their first term while 64.45% (n = 317,949) completed their first term. For officers, 3.43% (n = 90) attrited in their first term while 96.57% (n = 2,537) completed their first term.

Categorical Predictors. For a summary of results, see Exhibit 31. Rank Group showed to have a large effect and Soldier Sex had a small effect. For Rank Group, enlisted Soldiers (35.55%) were more likely to attrit than officers (3.42%), *OR* = 15.55, 95% CI [12.60, 19.18]. For Soldier Sex, female Soldiers (44.99%) were more likely to attrit than male Soldiers (33.67%), *OR* = 1.61, 95% CI [1.59, 1.64]. Similar patterns were found when looking at only enlisted Soldiers but not for officers.

Continuous Predictors. For a summary of results, see Exhibit 32. No relationships emerged at or above the small effect criteria when considering enlisted Soldiers and officers together. This pattern was the same when looking at enlisted Soldiers and officers separately.

Exhibit 32.

Categorical Predictors with First-Term Attrition by Sample Type

Predictors	*n*	*z*	*OR*	95 % CI	*p*	# of Event | Percentage of Event
Enlisted + Officer						
Rank Group	495,932	25.59	**15.55**	[12.60, 19.18]	< .001*	*Enlisted*: 175,356 (A), 317,949 (N) | 35.55% *Officer*: 90 (A), 2,537 (N) | 3.42%
MOS Type	495,932	*F* = 995.37	1.44	[1.42, 1.44]	< .001*	*Combat Arms*: 83,104 (A), 153,765 (N) | 35.08% *Combat Support*: 26,900 (A), 31,389 (N) | 46.15% *Combat Service Support*: 41,460 (A), 83,199 (N) | 33.26% *Special Service*: 11,855 (A), 33,172 (N) | 26.33% *Operations*: 234 (A), 336 (N) | 41.05% *Unknown*: 11,893 (A), 18,625 (N) | 38.97%
Soldier Sex	495,932	59.37	**1.61**	[1.59, 1.64]	< .001*	*Male*: 141,747 (A), 279,282 (N) | 33.67% *Female*: 33,699 (A), 41,204 (N) | 44.99%
Soldier Race	495,932	*F* = 85.17	1.12	[1.10, 1.12]	< .001*	*White*: 130,062 (A), 232,136 (N) | 35.91% *Black*: 36,118 (A), 68,308 (N) | 34.59% *Asian*: 6,984 (A), 15,314 (N) | 31.32% *American Indian/Alaskan Native*: 1,516 (A), 2,512 (N) | 37.64% *Native Hawaiian/Pacific Islander*: 592 (A), 1,407 (N) | 29.61% *Mixed Race/Other*: 174 (A), 809 (N) | 17.70%
Enlisted						
Rank Group	—	—	—	—	—	—
MOS Type	493,305	*F* = 1,023.60	1.44	[1.44, 1.47]	< .001*	*Combat Arms*: 83,061 (A), 152,702 (N) | 35.23% *Combat Support*: 26,884 (A), 30,753 (N) | 46.64% *Combat Service Support*: 41,434 (A), 82,534 (N) | 33.42% *Special Service*: 11,850 (A), 33,013 (N) | 26.41% *Operations*: 234 (A), 335 (N) | 41.12% *Unknown*: 11,893 (A), 18,612 (N) | 38.99%
Soldier Sex	493,305	59.41	**1.61**	[1.59, 1.64]	< .001*	*Male*: 141,670 (A), 277,112 (N) | 33.83% *Female*: 33,686 (A), 40,837 (N) | 45.20%
Soldier Race	493,305	*F* = 84.38	1.11	[1.10, 1.12]	< .001*	*White*: 129,990 (A), 230,116 (N) | 36.10% *Black*: 36,107 (A), 67,991 (N) | 34.69% *Asian*: 6,981 (A), 15,156 (N) | 31.54% *American Indian/Alaskan Native*: 1,516 (A), 2,503 (N) | 37.72% *Native Hawaiian/Pacific Islander*: 592 (A), 1,407 (N) | 29.61% *Mixed Race/Other*: 170 (A), 776 (N) | 17.97%
Officer						
Rank Group	—	—	—	—	—	—
MOS Type	2,627	*F* = 0.80	1.15	[1.00, 1.22]	.546	*Combat Arms*: 43 (A), 1,063 (N) | 3.89% *Combat Support*: 16 (A), 636 (N) | 2.45% *Combat Service Support*: 26 (A), 665 (N) | 3.76% *Special Service*: 5 (A), 159 (N) | 3.05% *Operations*: 0 (A), 1 (N) | 0.00% *Unknown*: 0 (A), 13 (N) | 0.00%
Soldier Sex	2,627	0.01	1.00	[0.57, 1.91]	.995	*Male*: 77 (A), 2,170 (N) | 3.43% *Female*: 13 (A), 367 (N) | 3.42%
Soldier Race	2,627	*F* = 1.50	1.21	[1.00, 1.30]	.200	*White*: 72 (A), 2,020 (N) | 3.44% *Black*: 11 (A), 317 (N) | 3.35% *Asian*: 3 (A), 158 (N) | 1.86% *American Indian/Alaskan Native*: 0 (A), 9 (N) | 0.00% *Native Hawaiian/Pacific Islander*: NC *Mixed Race/Other*: 4 (A), 33 (N) | 3.44%

Note. *n* = sample size, *F* = *F*-test coefficient, *OR* = odds ratio, CI = confidence interval of effect size, *p* = p-value, A = attrit, N = no attrit, NC = no cases, **p* < .05 after applying family-wise Bonferroni correction (α/882 = 0.000057); effect size values in bold represent effect sizes greater than trivial according to Chen et al. (2010).

Exhibit 33.

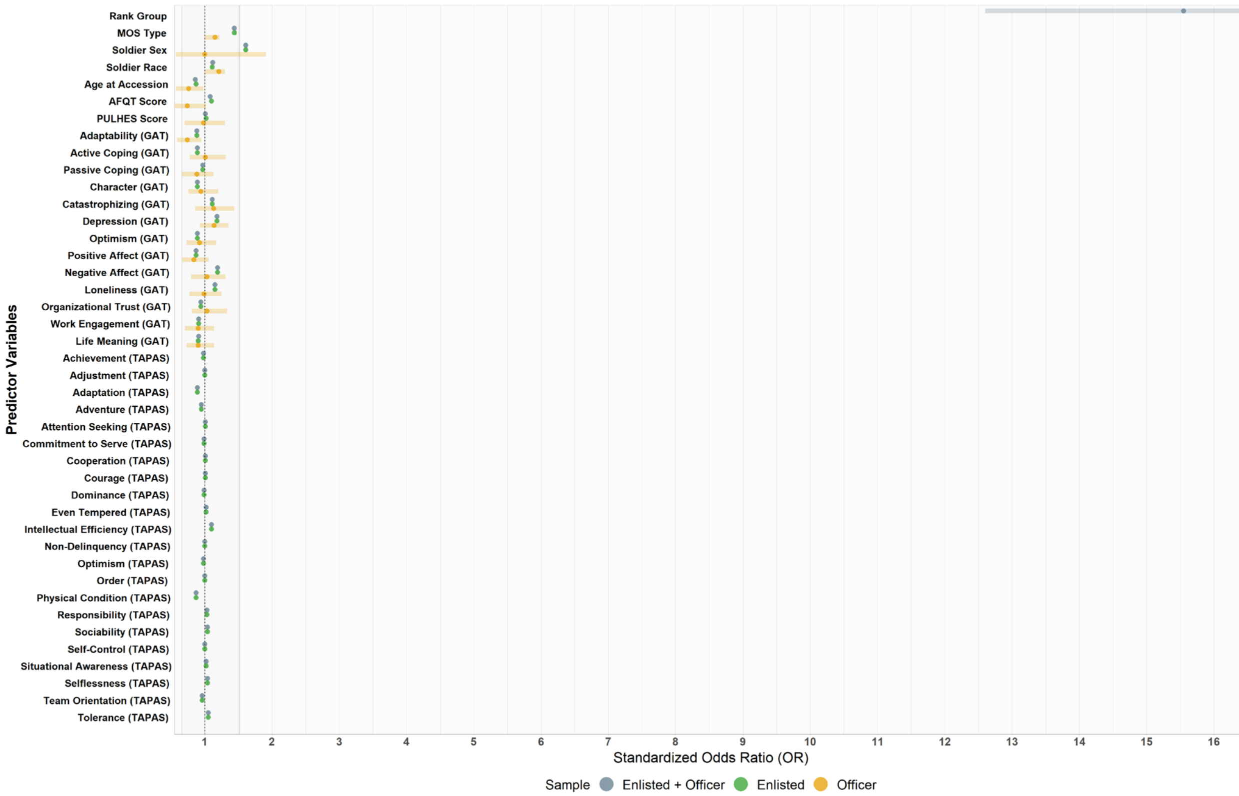
Continuous Predictors with First-Term Attrition by Sample Type

	Enlisted + Officer						*Enlisted*						*Officer*					
Predictors	***n***	***b***	**95% CI**	***β***	***OR***	***p***	***n***	***b***	**95% CI**	***β***	***OR***	***p***	***n***	***b***	**95% CI**	***β***	***OR***	***p***
Accession																		
Age at Accession	495,932	−0.04	[−0.04, −0.03]	−0.15	0.86	< .001*	493,305	−0.03	[−0.04, −0.03]	−0.14	0.87	< .001*	2,627	−0.06	[−0.11, 0.00]	−0.28	0.76	.046
AFQT Score	495,932	0.00	[0.00, 0.00]	0.08	1.08	< .001*	493,305	0.00	[0.00, 0.01]	0.09	1.10	< .001*	2,627	−0.02	[−0.04, 0.00]	−0.30	0.74	.058
PULHES Score	486,899	0.11	[0.07, 0.16]	0.01	1.01	< .001*	484,393	0.12	[0.07, 0.16]	0.02	1.02	< .001*	2,506	−0.13	[−1.84, 1.39]	−0.02	0.98	.879
GAT																		
Adaptability	323,053	−0.17	[−0.18, −0.16]	−0.13	0.88	< .001*	321,197	−0.16	[−0.17, −0.16]	−0.13	0.88	< .001*	1,856	−0.44	[−0.79, −0.08]	−0.30	0.74	.014
Active Coping	323,044	−0.16	[−0.17, −0.15]	−0.12	0.89	< .001*	321,188	−0.16	[−0.17, −0.15]	−0.12	0.89	< .001*	1,856	0.01	[−0.40, 0.44]	0.01	1.01	.949
Passive Coping	323,044	−0.03	[−0.04, −0.02]	−0.03	0.97	< .001*	321,188	−0.03	[−0.04, −0.02]	−0.03	0.97	< .001*	1,856	−0.15	[−0.48, 0.17]	−0.13	0.88	.359
Character	323,046	−0.08	[−0.08, −0.07]	−0.11	0.89	< .001*	321,190	−0.07	[−0.08, −0.07]	−0.11	0.89	< .001*	1,856	−0.05	[−0.24, 0.15]	−0.06	0.94	.625
Catastrophizing	323,053	0.12	[0.12, 0.13]	0.10	1.11	< .001*	321,197	0.12	[0.11, 0.13]	0.10	1.11	< .001*	1,856	0.17	[−0.22, 0.54]	0.12	1.13	.364
Depression	323,048	0.21	[0.20, 0.22]	0.16	1.18	< .001*	321,192	0.20	[0.20, 0.21]	0.16	1.18	< .001*	1,856	0.25	[−0.14, 0.57]	0.13	1.14	.172
Optimism	310,968	−0.15	[−0.16, −0.14]	−0.12	0.89	< .001*	309,162	−0.15	[−0.16, −0.14]	−0.12	0.89	< .001*	1,806	−0.12	[−0.45, 0.22]	−0.09	0.92	.475
Positive Affect	310,966	−0.17	[−0.18, −0.16]	−0.14	0.87	< .001*	309,160	−0.17	[−0.18, −0.16]	−0.14	0.87	< .001*	1,806	−0.26	[−0.60, 0.09]	−0.17	0.84	.146
Negative Affect	310,966	0.24	[0.23, 0.25]	0.17	1.19	< .001*	309,160	0.24	[0.23, 0.25]	0.18	1.19	< .001*	1,806	0.05	[−0.40, 0.47]	0.03	1.03	.822
Loneliness	323,053	0.16	[0.15, 0.17]	0.14	1.15	< .001*	321,197	0.16	[0.15, 0.17]	0.14	1.15	< .001*	1,856	−0.02	[−0.36, 0.30]	−0.02	0.99	.903
Organizational Trust	323,054	−0.08	[−0.09, −0.07]	−0.06	0.94	< .001*	321,198	−0.08	[−0.09, −0.07]	−0.06	0.94	< .001*	1,856	0.04	[−0.31, 0.41]	0.03	1.03	.824
Work Engagement	206,338	−0.10	[−0.11, −0.09]	−0.10	0.91	< .001*	204,730	−0.10	[−0.11, −0.09]	−0.10	0.91	< .001*	1,608	−0.14	[−0.45, 0.18]	−0.11	0.90	.367
Life Meaning	323,045	−0.11	[−0.12, −0.11]	−0.10	0.91	< .001*	321,189	−0.11	[−0.12, −0.11]	−0.10	0.90	< .001*	1,856	−0.13	[−0.41, 0.17]	−0.10	0.90	.381
TAPAS																		
Achievement	179,465	−0.05	[−0.07, −0.03]	−0.02	0.98	< .001*	179,451	−0.05	[−0.07, −0.03]	−0.02	0.98	< .001*	14	ISS	ISS	ISS	ISS	ISS
Adjustment	122,622	0.01	[−0.02, 0.03]	0.00	1.00	.470	122,609	0.01	[−0.02, 0.03]	0.00	1.00	.464	13	ISS	ISS	ISS	ISS	ISS
Adaptation	179,565	−0.12	[−0.13, −0.11]	−0.12	0.89	< .001*	179,451	−0.12	[−0.13, −0.11]	−0.12	0.89	< .001*	14	ISS	ISS	ISS	ISS	ISS
Adventure	29,659	−0.09	[−0.13, −0.05]	−0.05	0.95	< .001*	29,659	−0.09	[−0.13, −0.05]	−0.05	0.95	< .001*	0	ISS	ISS	ISS	ISS	ISS
Attention Seeking	141,620	0.03	[0.01, 0.04]	0.01	1.01	.009	141,607	0.03	[0.01, 0.04]	0.01	1.01	.009	13	ISS	ISS	ISS	ISS	ISS
Commitment to Serve	67,682	−0.03	[−0.06, 0.00]	−0.01	0.99	.060	67,681	−0.03	[−0.06, 0.00]	−0.01	0.99	.059	0	ISS	ISS	ISS	ISS	ISS
Cooperation	112,066	0.02	[−0.01, 0.04]	0.01	1.01	.158	112,056	0.02	[−0.01, 0.04]	0.01	1.01	.161	10	ISS	ISS	ISS	ISS	ISS
Courage	48,401	0.02	[−0.01, 0.06]	0.01	1.01	.232	48,397	0.02	[−0.01, 0.06]	0.01	1.01	.227	4	ISS	ISS	ISS	ISS	ISS
Dominance	179,465	−0.02	[−0.04, 0.00]	−0.01	0.99	.024	179,451	−0.02	[−0.04, 0.00]	−0.01	0.99	.025	14	ISS	ISS	ISS	ISS	ISS
Even Tempered	179,465	0.05	[0.03, 0.06]	0.02	1.02	< .001*	179,451	0.05	[0.03, 0.06]	0.02	1.02	< .001*	14	ISS	ISS	ISS	ISS	ISS
Intellectual Efficiency	179,465	0.17	[0.15, 0.19]	0.09	1.10	< .001*	179,451	0.17	[0.15, 0.19]	0.09	1.10	< .001*	14	ISS	ISS	ISS	ISS	ISS
Non-Delinquency	141,442	0.00	[−0.02, 0.02]	0.00	1.00	.855	141,429	0.00	[−0.02, 0.02]	0.00	1.00	.861	13	ISS	ISS	ISS	ISS	ISS
Optimism	179,465	−0.05	[−0.07, −0.03]	−0.02	0.98	< .001*	179,451	−0.05	[−0.07, −0.03]	−0.02	0.98	< .001*	14	ISS	ISS	ISS	ISS	ISS
Order	149,884	−0.01	[−0.03, 0.01]	−0.00	1.00	.471	149,874	−0.01	[−0.03, 0.01]	−0.00	1.00	.474	10	ISS	ISS	ISS	ISS	ISS
Physical Condition	179,465	−0.24	[−0.25, −0.22]	−0.14	0.87	< .001*	179,451	−0.24	[−0.25, −0.22]	−0.14	0.87	< .001*	14	ISS	ISS	ISS	ISS	ISS
Responsibility	48,606	0.05	[0.01, 0.09]	0.03	1.03	.007	48,601	0.05	[0.01, 0.09]	0.03	1.03	.007	5	ISS	ISS	ISS	ISS	ISS
Sociability	149,806	0.07	[0.05, 0.09]	0.04	1.04	< .001*	149,792	0.07	[0.05, 0.09]	0.04	1.04	< .001*	14	ISS	ISS	ISS	ISS	ISS
Self-Control	92,963	0.01	[−0.02, 0.03]	0.00	1.00	.498	92,950	0.01	[−0.02, 0.03]	0.00	1.00	.512	13	ISS	ISS	ISS	ISS	ISS
Situational Awareness	48,479	0.03	[0.00, 0.07]	0.02	1.02	.068	48,479	0.03	[0.00, 0.07]	0.02	1.02	.068	0	ISS	ISS	ISS	ISS	ISS
Selflessness	149,884	0.08	[0.06, 0.11]	0.04	1.04	< .001*	149,874	0.08	[0.06, 0.11]	0.04	1.04	< .001*	10	ISS	ISS	ISS	ISS	ISS
Team Orientation	48,579	−0.07	[−0.11, −0.04]	−0.04	0.96	< .001	48,575	−0.07	[−0.11, −0.04]	−0.04	0.96	< .001	0	ISS	ISS	ISS	ISS	ISS
Tolerance	149,806	0.09	[0.07, 0.11]	0.05	1.05	< .001*	149,792	0.09	[0.07, 0.11]	0.05	1.05	< .001*	14	ISS	ISS	ISS	ISS	ISS

Note. GAT = Global Assessment Tool, TAPAS = Tailor Adaptive Personality Assessment System, n = sample size, b = log odds ratio coefficient, CI = confidence interval, β = standardized log odds ratio coefficient, OR = odds ratio, p = p-value, *p < .05 after applying family-wise Bonferroni correction (α/882 = 0.000057), ISS = Inadequate sample size (n < 35); OR values in bold represent effect sizes greater than trivial according to Chen et al. (2010).

Exhibit 34.

Plot of Simple Relationship Strength between Predictor Variables and First-Term Attrition



Note. Shaded area represents trivial effect size according to Chen et al. (2010). Scale is truncated above 16.5.

Character of Service

To examine the simple relationships of Character of Service, a series of generalized linear models with a logit linkage were used for both categorical predictors and continuous predictors. A summary of standardized effect sizes for each predictor is presented in Exhibit 36. Overall, 1.68% (*n* = 7,164) separated under not honorable conditions while 98.32% (*n* = 418,902) separated with honorable discharges. For enlisted Soldiers, 1.72% (*n* = 7,094) separated under not honorable conditions while 98.28% (*n* = 405,406) separated with honorable discharges. For officers, 0.52% (*n* = 70) separated under not honorable conditions while 99.48% (*n* = 13,496) separated with honorable discharges.

Categorical Predictors. For a summary of results, see Exhibit 34. Rank Group showed to have a medium effect and Soldier Sex had a small effect. For Rank Group, enlisted Soldiers (1.72%) were more likely to have not honorable discharges than officers (0.52%), *OR* = 3.37, 95% CI [2.66, 4.27]. For Soldier Sex, male Soldiers (1.85%) were more likely to have not honorable discharges than female Soldiers (0.71%), *OR* = 2.63, 95% CI [2.39, 2.90]. Similar patterns were found when looking at only enlisted Soldiers and officers.

Continuous Predictors. For a summary of results, see Exhibit 35. No relationships emerged at or above the small effect criteria when considering enlisted Soldiers and officers together. This pattern was the same when only looking at enlisted Soldiers. For officers, medium to small effects emerged. Positive effects included Achievement (+), Order (+), Selflessness (+), and Tolerance (+). Negative effects included Adaptation (−), Cooperation (−), Dominance (−), Optimism (TAPAS; −), Physical Condition (−), and Self-Control (−), Officer effects were modeled with relatively low sample sizes and, in some cases, were insufficient to obtain estimates, so inferences should be made with caution.

Exhibit 35.

Categorical Predictors with Character of Service by Sample Type

Predictors	*n*	*z*	*OR*	95 % CI	*p*	# of Event | Percentage of Event
Enlisted + Officer						
Rank Group	426,066	10.10	**3.37**	[2.66, 4.27]	< .001*	*Enlisted*: 7,094 (NH), 405,406 (H) | 1.72% *Officer*: 70 (NH), 13,496 (H) | 0.52%
MOS Type	426,031	*F* = 5.31	1.03	[1.02, 1.04]	< .001	*Combat Arms*: 3,629 (NH), 201,755 (H) | 1.77% *Combat Support*: 841 (NH), 47,428 (H) | 1.74% *Combat Service Support*: 1,689 (NH), 104,918 (H) | 1.58% *Special Service*: 664 (NH), 42,138 (H) | 1.55% *Operations*: 8 (NH), 542 (H) | 1.45% *Unknown*: 332 (NH), 22,087 (H) | 1.48%
Soldier Sex	426,066	19.74	**2.63**	[2.39, 2.90]	< .001*	*Male*: 6,716 (NH), 356,378 (H) | 1.85% *Female*: 448 (NH), 62,524 (H) | 0.71%
Soldier Race	426,066	*F* = 49.81	1.09	[1.09, 1.10]	< .001*	*White*: 4,889 (NH), 308,175 (H) | 1.56% *Black*: 1,960 (NH), 85,407 (H) | 2.24% *Asian*: 194 (NH), 18,666 (H) | 1.03% *American Indian/Alaskan Native*: 78 (NH), 3,378 (H) | 2.26% *Native Hawaiian/Pacific Islander*: 13 (NH), 1,493 (H) | 0.86% *Mixed Race/Other*: 30 (NH), 1,783 (H) | 1.65%
Enlisted						
Rank Group	—	—	—	—	—	—
MOS Type	412,500	*F* = 5.38	1.03	[1.02, 1.04]	< .001	*Combat Arms*: 3,604 (NH), 196,012 (H) | 1.81% *Combat Support*: 830 (NH), 45,972 (H) | 1.77% *Combat Service Support*: 1,681 (NH), 102,628 (H) | 1.61% *Special Service*: 642 (NH), 38,253 (H) | 1.65% *Operations*: 6 (NH), 519 (H) | 1.14% *Unknown*: 331 (NH), 22,022 (H) | 1.48%
Soldier Sex	412,500	19.19	**2.58**	[2.34, 2.84]	< .001*	*Male*: 6,653 (NH), 346,225 (H) | 1.89% *Female*: 441 (NH), 59,181 (H) | 0.74%
Soldier Race	412,500	*F* = 45.93	1.09	[1.08, 1.10]	< .001*	*White*: 4,840 (NH), 297,329 (H) | 1.60% *Black*: 1,948 (NH), 84,502 (H) | 2.25% *Asian*: 190 (NH), 17,806 (H) | 1.06% *American Indian/Alaskan Native*: 78 (NH), 3,311 (H) | 2.30% *Native Hawaiian/Pacific Islander*: 13 (NH), 1,481 (H) | 0.87% *Mixed Race/Other*: 25 (NH), 977 (H) | 2.50%
Officer						
Rank Group	—	—	—	—	—	—
MOS Type	13,531	*F* = 2.39	1.11	[1.02, 1.15]	.035	*Combat Arms*: 25 (NH), 5,743 (H) | 0.43% *Combat Support*: 11 (NH), 1,456 (H) | 0.75% *Combat Service Support*: 8 (NH). 2,290 (H) | 0.35% *Special Service*: 22 (NH), 3,885 (H) | 0.56% *Operations*: 2 (NH), 23 (H) | 8.00% *Unknown*: 1 (NH), 65 (H) | 1.52%
Soldier Sex	13,566	2.72	**2.96**	[1.45, 7.12]	.006	*Male*: 63 (NH), 10,153 (H) | 0.62% *Female*: 7 (NH), 3,343 (H) | 0.21%
Soldier Race	13,566	*F* = 1.97	1.10	[1.00, 1.14]	.080	*White*: 49 (NH), 10,846 (H) | 0.45% *Black*: 12 (NH), 905 (H) | 1.31% *Asian*: 4 (NH), 860 (H) | 0.46% *American Indian/Alaskan Native*: 0 (NH), 67 (H) | 0.00% *Native Hawaiian/Pacific Islander*: 0 (NH), 12 (H) | 0.00% *Mixed Race/Other*: 5 (NH), 806 (H) | 0.62%

Note. *n* = sample size, *F* = *F*-test coefficient, *OR* = odds ratio, CI = confidence interval of effect size, *p* = p-value, **p* < .05 after applying family-wise Bonferroni correction (α/882 = 0.000057), NH = not honorable discharge, H = honorable discharge, NC = no cases; effect size values in bold represent effect sizes greater than trivial according to Chen et al. (2010).

Exhibit 36.

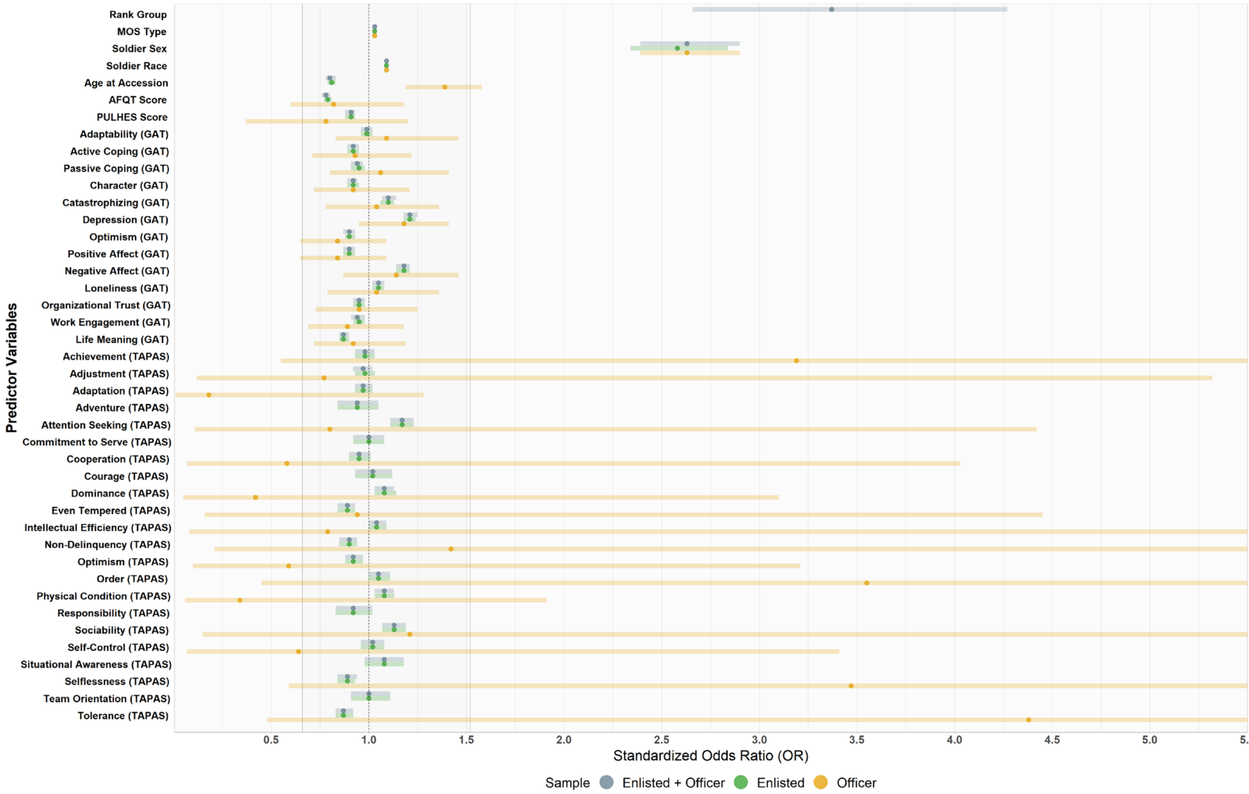
Continuous Predictors with Character of Service by Sample Type

Predictors	*Enlisted + Officer*						*Enlisted*						*Officer*					
	n	***b***	**95% CI**	***β***	***OR***	***p***	***n***	***b***	**95% CI**	***β***	***OR***	***p***	***n***	***b***	**95% CI**	***β***	***OR***	***p***
Accession																		
Age at Accession	426,066	−0.05	[−0.06, −0.05]	−0.22	0.80	< .001*	412,500	−0.05	[−0.06, −0.04]	−0.21	0.81	< .001*	13,566	0.07	[0.04, 0.09]	0.33	1.39	< .001*
AFQT Score	416,221	−0.01	[−0.01, −0.01]	−0.25	0.78	< .001*	412,174	−0.01	[−0.01, −0.01]	−0.24	0.79	< .001*	4,047	−0.01	[−0.03, 0.01]	−0.20	0.82	.255
PULHES Score	408,080	−0.74	[−0.95, −0.53]	−0.10	0.91	< .001*	404,573	−0.74	[−0.95, −0.53]	−0.10	0.91	< .001*	3,507	−1.30	[−5.20, 0.96]	−0.25	0.78	.428
GAT																		
Adaptability	285,635	−0.01	[−0.05, 0.03]	−0.01	0.99	.580	276,071	−0.01	[−0.05, 0.03]	−0.01	0.99	.587	9,564	0.13	[−0.28, 0.57]	0.09	1.09	.538
Active Coping	285,626	−0.12	[−0.16, −0.07]	−0.09	0.92	< .001*	276,062	−0.11	[−0.15, −0.07]	−0.09	0.92	< .001*	9,564	−0.13	[−0.56, 0.33]	−0.08	0.93	.577
Passive Coping	285,626	−0.07	[−0.10, −0.03]	−0.06	0.94	< .001	276,062	−0.06	[−0.10, −0.03]	−0.06	0.95	< .001	9,564	0.07	[−0.26, 0.40]	0.06	1.06	.674
Character	285,627	−0.06	[−0.08, −0.04]	−0.09	0.92	< .001*	276,066	−0.06	[−0.08, −0.04]	−0.09	0.92	< .001*	9,561	−0.07	[−0.27, 0.16]	−0.08	0.92	.545
Catastrophizing	285,637	0.12	[0.08, 0.15]	0.10	1.10	< .001*	276,073	0.11	[0.08, 0.15]	0.09	1.10	< .001*	9,564	0.06	[−0.36, 0.44]	0.04	1.04	.759
Depression	285,633	0.25	[0.21, 0.28]	0.19	1.21	< .001*	276,069	0.24	[0.20, 0.27]	0.19	1.21	< .001*	9,564	0.32	[−0.09, 0.65]	0.17	1.18	.094
Optimism	276,528	−0.14	[−0.18, −0.10]	−0.11	0.90	< .001*	267,140	−0.13	[−0.17, −0.09]	−0.10	0.90	< .001*	9,388	−0.26	[−0.63, 0.12]	−0.18	0.84	.176
Positive Affect	276,527	−0.14	[−0.17, −0.10]	−0.11	0.90	< .001*	267,139	−0.13	[−0.17, −0.09]	−0.11	0.90	< .001*	9,388	−0.27	[−0.65, 0.13]	−0.18	0.84	.179
Negative Affect	276,527	0.23	[0.18, 0.27]	0.16	1.18	< .001*	267,139	0.22	[0.18, 0.27]	0.16	1.18	< .001*	9,388	0.23	[−0.24, 0.67]	0.13	1.14	.323
Loneliness	285,637	0.06	[0.02, 0.09]	0.05	1.05	.002	276,073	0.05	[0.02, 0.09]	0.05	1.05	.003	9,564	0.06	[−0.32, 0.42]	0.04	1.04	.763
Organizational Trust	285,639	−0.07	[−0.10, −0.03]	−0.05	0.95	.001	276,075	−0.07	[−0.10, −0.03]	−0.05	0.95	< .001	9,564	−0.07	[−0.45, 0.33]	−0.05	0.95	.709
Work Engagement	195,730	−0.06	[−0.10, −0.02]	−0.06	0.94	.001	187,726	−0.05	[−0.09, −0.02]	−0.05	0.95	.004	8,004	−0.15	[−0.49, 0.22]	−0.11	0.89	.405
Life Meaning	285,628	−0.16	[−0.19, −0.12]	−0.14	0.87	< .001*	276,064	−0.16	[−0.19, −0.12]	−0.14	0.87	< .001*	9,564	−0.11	[−0.42, 0.23]	−0.08	0.92	.508
TAPAS																		
Achievement	141,726	−0.04	[−0.14, 0.05]	−0.02	0.98	.369	141,526	−0.05	[−0.14, 0.05]	−0.02	0.98	.355	200	2.38	[−1.21, 6.72]	1.16	**3.19**	.182
Adjustment	102,348	−0.06	[−0.18, 0.05]	−0.03	0.97	.272	102,151	−0.06	[−0.18, 0.05]	−0.02	0.98	.272	197	−0.49	[−4.16, 3.22]	−0.26	0.77	.798
Adaptation	141,726	−0.03	[−0.08, 0.02]	−0.03	0.97	.256	141,526	−0.03	[−0.07, 0.02]	−0.03	0.97	.279	200	−1.86	[−4.56, 0.27]	−1.72	**0.18**	.099
Adventure	24,101	−0.10	[−0.29, 0.08]	−0.06	0.94	.285	24,079	−0.10	[−0.29, 0.09]	−0.06	0.94	.287	22	ISS	ISS	ISS	ISS	ISS
Attention Seeking	115,646	0.27	[0.18, 0.36]	0.15	1.17	< .001*	115,448	0.27	[0.18, 0.36]	0.15	1.17	< .001*	198	−0.38	[−3.80, 2.59]	−0.22	0.80	.829
Commitment to Serve	50,474	−0.00	[−0.16, 0.15]	−0.00	1.00	.961	50,449	−0.00	[−0.16, 0.15]	−0.00	1.00	.962	25	ISS	ISS	ISS	ISS	ISS
Cooperation	91,515	−0.10	[−0.22, 0.02]	−0.05	0.95	.101	91,344	−0.10	[−0.22, 0.02]	−0.05	0.95	.102	171	−1.28	[−6.43, 3.33]	−0.54	**0.58**	.600
Courage	36,913	0.03	[−0.14, 0.21]	0.02	1.02	.706	36,885	0.03	[−0.14, 0.21]	0.02	1.02	.704	28	ISS	ISS	ISS	ISS	ISS
Dominance	141,726	0.14	[0.06, 0.23]	0.08	1.08	.001	141,526	0.15	[0.06, 0.23]	0.08	1.08	< .001	200	−1.64	[−5.80, 2.12]	−0.87	**0.42**	.391
Even Tempered	141,726	−0.23	[−0.32, −0.14]	−0.12	0.89	< .001*	141,526	−0.23	[−0.32, −0.14]	−0.12	0.89	< .001*	200	−0.14	[−3.85, 3.19]	−0.07	0.94	.941
Intellectual Efficiency	141,726	0.08	[−0.01, 0.17]	0.04	1.04	.080	141,526	0.08	[−0.01, 0.17]	0.04	1.04	.077	200	−0.43	[−4.59, 3.43]	−0.24	0.79	.825
Non-Delinquency	115,353	−0.22	[−0.32, −0.11]	−0.11	0.90	< .001*	115,156	−0.22	[−0.32, −0.11]	−0.11	0.90	< .001*	197	0.72	[−3.19, 4.54]	0.35	1.42	.739
Optimism	141,726	−0.19	[−0.29, −0.08]	−0.08	0.92	.001	141,526	−0.18	[−0.29, −0.08]	−0.08	0.92	< .001	200	−1.14	[−5.01, 2.52]	−0.53	**0.59**	.579
Order	117,818	0.09	[0.00, 0.19]	0.05	1.05	.046	117,646	0.09	[0.00, 0.19]	0.05	1.05	.050	172	2.11	[−1.34, 6.89]	1.27	**3.55**	.295
Physical Condition	141,726	0.13	[0.05, 0.21]	0.08	1.08	.001	141,526	0.13	[0.05, 0.21]	0.08	1.08	.001	200	−1.90	[−4.77, 1.13]	−1.09	**0.34**	.165
Responsibility	36,983	−0.17	[−0.38, 0.03]	−0.08	0.92	.102	36,953	−0.17	[−0.38, 0.03]	−0.08	0.92	.104	30	ISS	ISS	ISS	ISS	ISS
Sociability	117,625	0.21	[0.12, 0.29]	0.12	1.13	< .001*	117,447	0.21	[0.12, 0.29]	0.12	1.13	< .001*	178	0.33	[−3.19, 4.20]	0.19	1.21	.855
Self-Control	78,247	0.04	[−0.08, 0.15]	0.02	1.02	.538	78,072	0.04	[−0.08, 0.16]	0.02	1.02	.525	175	−0.85	[−5.09, 2.36]	−0.44	**0.64**	.659
Situational Awareness	37,106	0.15	[−0.04, 0.34]	0.07	1.08	.113	37,084	0.15	[−0.04, 0.34]	0.07	1.08	.113	22	ISS	ISS	ISS	ISS	ISS
Selflessness	117,818	−0.25	[−0.37, −0.14]	−0.12	0.89	< .001*	117,646	−0.26	[−0.37, −0.15]	−0.12	0.89	< .001*	172	2.66	[−1.11, 6.38]	1.24	**3.47**	.118
Team Orientation	37,206	0.01	[−0.19, 0.20]	0.00	1.00	.955	37,177	0.01	[−0.19, 0.20]	0.00	1.00	.959	29	ISS	ISS	ISS	ISS	ISS
Tolerance	117,625	−0.24	[−0.34, −0.15]	−0.14	0.87	< .001*	117,447	−0.25	[−0.34, −0.15]	−0.14	0.87	< .001*	178	2.53	[−1.25, 7.77]	1.48	**4.38**	.257

Note. GAT = Global Assessment Tool, TAPAS = Tailor Adaptive Personality Assessment System, *n* = sample size, *b* = log odds ratio coefficient, CI = confidence interval, *β* = standardized log odds ratio coefficient, *OR* = odds ratio, *p* = p-value, **p* < .05 after applying family-wise Bonferroni correction (α/882 = 0.000057), ISS = Inadequate sample size (*n* < 35); *OR* values in bold represent effect sizes greater than trivial according to Chen et al. (2010).

Exhibit 37.

Plot of Simple Relationship Strength between Predictor Variables and Character of Service



Note. Shaded area represents trivial effect size according to Chen et al. (2010). Scale truncated at 5.5.

Discussion

Summary of Findings

The goal of this research was to examine simple relationships between predictor variables and outcome variables related to individual work performance. Our efforts to conduct an initial analytical modeling of variables in the PDE were fourfold.

First, through a process of conceptual profiling we analyzed the complete corpus of data tables and variables available to us in the PDE (> 3,500 variables). Through the use of past research and our conceptual performance model, we identified variables that may have some relation with performance. Importantly, this process highlighted a lack of outcome variables related to work performance in the administrative data corpus. In our profiling, we identified seven variables that had some tangible connection to our conceptualization of Soldier performance:

(a) Alcohol Use Disorders Identification Test-Concise (AUDIT-C) Scores;
(b) Army Physical Fitness Test (APFT) Scores;
(c) Award Counts (good conduct);
(d) Bad Paper Counts (bad conduct—Article 15s, Letters of Reprimand, Court Martials);
(e) Speed of Promotion;

(f) First-Term Attrition; and,
(g) Character of Service.

Most of these variables were characterized under the task performance (i.e., APFT Score), contextual performance (i.e., AUDIT-C Score, Award Count), counterproductive performance (i.e., Bad Paper Count), or a general performance/outcome variable category (i.e., Speed of Promotion, First-Term Attrition, Character of Service). To the authors' knowledge there appear to be little to no administrative data on Soldier performance that would be documented in performance review forms, supervisor ratings, or other performance metrics assessing a Soldier's overall performance in an Army unit. This finding stands in contrast to prior research projects (e.g., Project A), which had access to many other non-administrative variables related to occupation (MOS)-specific tasks criteria, MOS-specific behaviors, and Army-wide rating scales (for a review, see Campbell & Zook, 1991). This evidence reveals major gaps in the data collected by the Army on performance. For better modeling of Soldier performance in the Army, it would be beneficial if the Army collected more detailed information on Soldiers (e.g., supervisor ratings) to be included with other administrative data. Even if the data were collected for simple research/feedback purposes (versus high-stakes promotion criteria), this information would be useful for enabling more nuanced models of individual performance in the Army.

Second, descriptive statistics were derived for the identified variables related to performance. Descriptive statistics provided insight into the central tendency, dispersion, and distribution of variables as well as the percent missing. A few notable findings emerged from this analysis including (a) the sample characteristics of Soldiers (e.g., Race, Sex, MOS Type, Age) was consistent with past Army demographic trends for the same time period; (b) most variables (both predictor and outcome) were within their expected ranges and approximated normal distributions respective of large sample sizes; (c) the base rate for undesirable outcomes related to performance like receiving bad papers (i.e., Article 15, Letter of Reprimand, Courts Martial) or an undesirable discharge classification (i.e., not honorable) were very low (< 2% of cases); (d) the TAPAS had the highest prevalence of missing data (> 94%), especially for officers; and (e) measures on the GAT seemed to have the highest levels of zero-order correlations with one another (*r*s from 0.01 to 0.69). Overall, the descriptive statistics provided a baseline for understanding the composition of variables used in modeling of simple relationships between predictors and outcomes.

Third, a simple relationship analysis (e.g., zero-order correlations, simple linear regressions, one-way ANOVAs) was conducted to model relationships between a single predictor and a single outcome related to performance. Inferential statistics and indices of standardized effect size revealed which simple relationships were strongest and in which direction. A few notable findings emerged from this analysis including (a) the variables AUDIT-C Score and APFT Score showed the most number of predictors with effect sizes meeting or exceeding the 'small' threshold as guided by Cohen (1988); (b) Soldier Sex and Rank Group seemed to be the most consistently strong categorical predictor variables; (c) Life Meaning and Adventure seemed to be the most consistently strong continuous predictor variables; (d) certain relationships (e.g., Soldier Sex and

Age at Accession with AUDIT-C Score, MOS Type and Age at Accession with APFT Score, and Soldier Sex with Bad Paper Count) seem to show differences across the rank groupings suggesting potential interaction effects; and (e) variable combinations with high missing data rates (e.g., examining officers with TAPAS predictors greater than 99% missing in some cases) still had adequate sample size to examine most simple relationships, however, more complex relationships might be difficult to assess with these sample groups and predictors. Overall, the simple relationship analysis provided a solid baseline assessment of simple relationships to inform the building of models.

Lastly, a dominance analysis was conducted to further understand which predictor variables had the strongest and most consistent relationship with the different outcome variables in the presence of other strong predictors. For each outcome variable of interest (e.g., AUDIT-C Score, APFT Score, Award Count) the Top 7 predictors were chosen based on their standardized effect size from the simple relationship modeling. Findings from the dominance analysis indicated that Rank Group, Soldier Sex, and Age at Accession showed general and complete dominance across more than one of the performance-related outcome variables. However, in most cases, complete dominance between predictor variables could not be determined given equally strong relationships with respective outcome variables. The complete summary of the dominance analysis results are available in Ratcliff et al. (2021c), previously submitted to ARI for review.

Limitations and Future Directions

Future research should address limitations of the current research to better predict performance. First, as mentioned during the conceptual profiling stage of our research, there are few administrative variables that either directly or indirectly assessed individual performance. This pattern is especially apparent when considering task performance or the quantity and quality of the work Soldiers are asked to perform. Given the lack of administrative data related to formal performance evaluations, there is little administrative data that can serve as a proxy for how well Soldiers are performing in their assigned jobs which has been an issue raised by other research (see Horowitz et al., 2019). Access to formal performance evaluations like Officer Evaluation Reports (OERs), Noncommissioned Officer Evaluation Reports (NCOERs), or, for junior enlisted Soldiers, Developmental Counseling Forms (DA 4856) would be a first step to addressing these gaps.

Another potential avenue for gaining performance evaluations would be to directly collect new data and link it with the administrative data already obtained. Similar to the methods used in Project A (cf. Campbell & Zook, 1991; Pulakos & Borman, 1986), going to peers and supervisors of Soldiers using Army-wide ratings of performance would provide a direct measurement of performance across multiple dimensions. Not only would this approach provide a more nuanced assessment of performance and greater criterion validity, collecting new data would also provide opportunities to gain adjacent insights about Soldiers not contained in the administrative data (e.g., organizational climate, unit cohesion, perceived motivation). Direct data collection can often be a tool that is uniquely able to target research questions and fill in gaps of administrative data. An alternative might be to require Evaluation Reports to provide qualitative analysis (as is currently done) and a quantitative assessment that aligns with the qualitative assessment.

Second, there was a limited time frame for which administrative data was available. For this research, we examined Soldiers who accessed from 2008 to 2015 with data on these Soldiers extending into 2019. Although this 11-year time frame would cover the majority of Soldiers who serve for one or two terms (typically four-year increments), it does not capture those who stay in the Army for a full career of 20-25 years. This time frame would cover only up to a midpoint in the career of those Soldiers. Future research should continually update relevant data to expand the time frame out to cover a larger period of time.

Third, the Speed of Promotion outcome variable was calculated using a Soldier's starting rank and the highest rank they achieved as of the last date of the administrative records. However, rank data was censored for higher ranks (i.e., senior enlisted = Sergeant First Class and higher; senior officers = Lieutenant Colonel and higher). This censoring makes it difficult to see the highest rank achieved by Soldiers and could reduce variance in the calculation of Speed of Promotion. Future research should seek to obtain uncensored rank data to fully understand promotion rates through the ranks.

Fourth, many of the simple relationships observed were either trivial or small in magnitude using standardized metrics of effect size. Only a few simple relationships reached thresholds of medium or large as proposed by Cohen (1988). However, it should be noted that Cohen's (1988) thresholds for effect size magnitude were somewhat subjective and not necessarily applicable to all circumstances. For example, in high stakes contexts (e.g., suicide research) an effect size lower than the small threshold could be quite meaningful. We chose to follow Cohen's guidelines in the absence of prior research or guidance by the Army on what constitutes a meaningful effect in the realm of performance criteria. Thus, certain effects that fail to reach the small threshold may be important in the eyes of Army leaders and practitioners.

Analysis 3: Logistic Model of Officer Promotion Time

We expand on analysis of promotion time by considering statistical modeling techniques that account for interactions between variables. As a proof-of-concept this is done in the context of one of our indicators of performance, promotion time among Army Officers. Promotion time presents a useful test case, because it an outcome that can easily expanded from a threshold or binary outcome (i.e., was a Soldier promoted) into stratified outcomes including categories (i.e., above-, in-, or below-the-zone) or outcomes approaching continuous measures (e.g., months or days between promotions). While this modeling does expand on the simple relationships analysis to account for simple interactions between variables, we did not examine whether interaction terms should be included.

Officer Sample Data

Army administrative data from the PDE was collected for Army Officers with at least one quarterly observation from 2001 to 2020, which yields a cohort of 107,440 Army Officers for this analysis. Of these officers, 73,440 had observed GAT survey responses. We found some demographic differences between Officers with and without an observed GAT survey, but these differences are minimal (Ratcliff et al., 2021a; Ratcliff et al., 2021b).

Variables were identified in the PDE to serve as predictors for promotion speed and are constructed according to the methods described above. Variables consisted of Rank, Sex, Race, Education, Officer selection source, and measures of psychosocial traits from the GAT. Methods of cleaning and transforming each variable used in the analysis also follow the same process as described in earlier analyses.

Speed of promotion is the performance-related outcome. Specifically, we examine factors that are associated with a higher likelihood of below-the-zone promotions for Army Officers. Distributions of promotion time to O4 (Major) and to O5 (Lieutenant Colonel) are shown in Exhibit 37. Below the zone promotions are defined according to Army regulations (PAM 600-3), and opportunities for below the zone promotions begin at the rank of Major. A promotion “in the zone” to O4 is defined as a total time in service from 9 to 11 years. Similarly, promotions in the zone to O5 are defined as an overall time in service from 15 to 17 years, reflecting an additional time in grade from 4 to 6 years.

Exhibit 38.

Average Time to Promotion (Years) to O4 (Major) and O5 (Lieutenant Colonel) for Army Officers

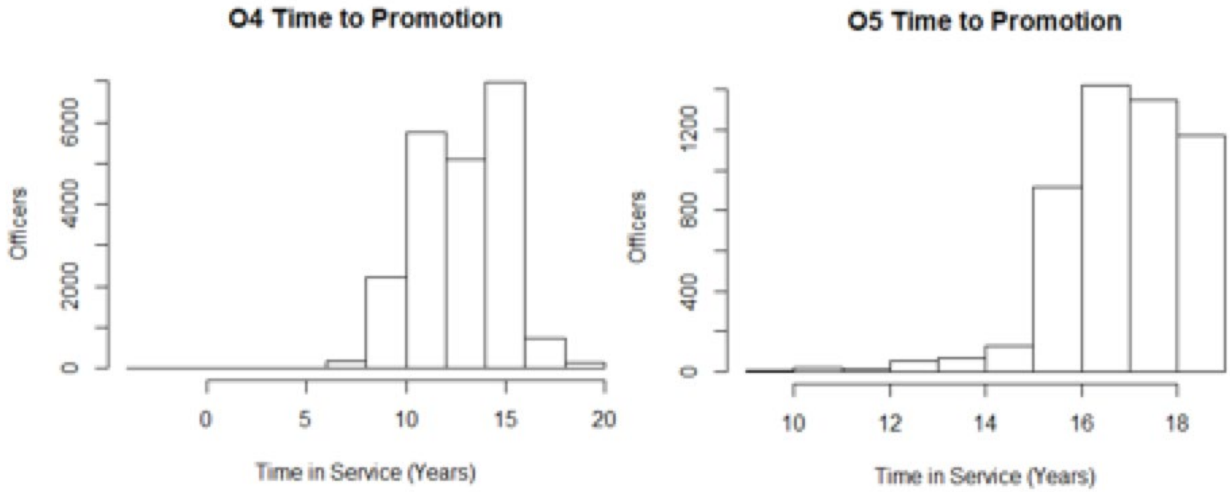


Exhibit 39.

Army Officer below the **zone and above the zone promotions.**

Rank	Reached rank by 2020	Below-the-zone (early)	In-the-zone	Above-the-zone (late)
Major (O4)	21,133	203	8,024	12,965
Lieutenant Colonel (O5)	5,150	160	4,846	144

Logistic Regression Model

The model we employed is as follows:

- Let y_i be the outcome variable of a below the zone promotion to O4;
- $y_i = 1$ when time to promotion for officer *i* is less than 11 years and $y_i = 0$ otherwise;
- Let $\boldsymbol{X}$ be the vector of Army administrative and GAT covariates; then we model the likelihood of below the zone promotion as $\text{logit}(y_i) = \boldsymbol{X\beta}$.

Results and coefficients are shown below.

Our model for below the zone promotion to O5 is similar; except that we add an additional covariate $b_i = 1$ if officer *i* had a below the zone promotion to O4. We add this to control for the effect that the overall number of early promotions that an officer can achieve is limited. Our model for below the zone promotions to O5 is $\text{logit}(y_i) = \boldsymbol{X\beta} + b_i$.

Results

Optimism and Work Engagement are psychosocial traits associated with a significantly increased likelihood of early promotions to both O4 and O5. Additionally, a higher educational background contributed to an increased chance of a below the zone promotion. For promotions to O5 we find that *b* is negatively significant, indicating that Officers who received an early promotion to Major are much less likely to receive another early promotion.

Exhibit 40.
Predictors of Below the Zone Promotion to O4 (Major)

Variable	Name	Estimate	SE	Z	p
Sex					
(ref: Female)	Male	-0.22	0.36	-0.6	0.54
Race					
(ref: White)	Black	-1.78	1.02	-1.83	0.08
	Asian	0.22	0.49	0.45	0.66
	Other	1.5	0.80	1.91	0.06
Marital Status					
(ref: Not married)	Married	-0.41	0.29	-1.41	0.16
Education					
(ref: High School)	Bachelors*	1.5	0.66	2.3	0.01*
	Doctorate	-1.4	1.2	-1.2	0.11
	Masters*	2.8	0.75	3.6	<.001*
	Professional	1.0	0.69	1.5	0.09
Selection Source					
(ref: ROTC)	Direct Commission	0.46	0.55	0.84	0.40
	OCS, OTS, or PLC	-1.69	0.83	-2.07	0.04*
	Military Academy*	0.99	0.30	3.3	<.001*
GAT Scales					
	Active Coping	0.49	0.29	1.6	0.10
	Passive Coping	0.21	0.16	1.3	0.21
	Character	-0.25	0.13	-1.8	0.07
	Catastrophizing	0.30	0.24	1.26	0.21
	Optimism*	0.5	0.25	1.98	0.04*
	Positive Affect	0.11	0.31	0.33	0.73
	Negative Affect	-0.49	0.30	-1.62	0.10
	Loneliness	-0.13	0.26	-0.51	0.61
	Organizational Trust	-0.21	0.21	-0.98	0.32
	Work Engagement	0.32	0.21	1.50	0.07
	Life Meaning	-0.02	0.21	-0.09	0.92

Exhibit 41.
Predictors of Below the Zone Promotion to O5 (Lieutenant Colonel)

Variable	Name	Estimate	SE	Z	p
Below the Zone O4 Promotion					
	b	-1.61	0.81	1.98	0.05*
Sex					
(ref: Female)	Male	-0.11	0.14	-0.81	0.41
Race					
(ref: White)	Black	0.12	0.14	0.81	0.41
	Asian	-0.07	0.22	-0.29	0.77
	Other	0.22	0.80	0.27	0.78
Marital Status					
(ref: Not married)	Married	-0.20	0.12	-1.65	0.10
Education					
(ref: High School)	Bachelors	0.24	0.72	0.3	0.38
	Doctorate	0.64	0.89	0.72	0.24
	Masters	0.17	0.73	0.23	0.42
	Professional	0.99	0.78	1.3	0.10
Selection Source					
(ref: ROTC)	Direct Commission*	1.63	0.65	2.48	0.01*
	OCS, OTS, or PLC	0.04	0.13	0.32	0.74
	Military Academy	0.23	0.12	1.82	0.07
GAT Scales					
	Active Coping	-0.01	0.10	-0.15	0.87
	Passive Coping	0.03	0.05	0.54	0.58
	Character	0.03	0.05	0.53	0.59
	Catastrophizing	0.11	0.10	1.04	0.30
	Optimism	0.25	0.13	1.8	0.07
	Positive Affect	0.10	0.11	0.88	0.38
	Negative Affect	0.10	0.11	0.81	0.41
	Loneliness	0.07	0.08	0.89	0.37
	Organizational Trust	0.06	0.08	0.74	0.46
	Work Engagement	0.20	0.07	2.49	0.01*
	Life Meaning	-0.11	0.08	-1.4	0.16

Discussion of Analysis 2 and 3

The findings from these analyses have broader implications for modeling performance in the U.S. Army using administrative data. First, this research demonstrates that the Army administrative data can be leveraged to model Soldier performance. This basic fact was taken as an unstated assumption at the outset of the project, but it was by no means certain. Though the process of linking and cleaning administrative data can be arduous, the administrative data provides insights into Soldiers that cover many different topics and over long periods of time. This enables researchers to glean an understanding of Soldiers that is difficult to capture through methods like surveys, which are limited in the information they can collect (e.g., due to survey burden or fatigue), costly, and limited in time (e.g., usually collected at a single time point).

In addition, this research informs current efforts in the Army related to talent management, selection and assignment, and Soldier effectiveness. By understanding the linkages between predictors and indicators of performance, the Army can better assess how to optimize performance capabilities, for example, selection into an occupation (MOS) and promotion decisions. The relationships described in this research mark a first step in unraveling the complicated interconnections between these variables and further modeling is needed to better understand how certain predictors are related and interact with other variables at predicting performance. With further modeling, the Army could develop an even more nuanced understanding of Soldier performance that takes into account even more dimensions of a Soldier and how these related to their effectiveness.

What is more, we demonstrate that performance indicators can be linked to an expanded view of performance outcomes in a meaningful way, revealing how aspects of the social dimension of Soldiers (chiefly the degree to which they maintain an optimistic outlook and the degree to which they feel organizational engaged) are linked to categorical outcomes that begin to capture a wider dimension of performance. As the more complex nature of these relationships are revealed, this will provide further understanding of individual performance to enable the Army to better utilize and manage personnel in building an effective and ready force.

In conclusion, this these analyses lay the foundation for modeling Soldier performance by identifying predictors and outcomes related to performance as well as providing modeling estimates between them. We propose that these relationships be examined in greater detail by incorporating a temporal dimension to the analyses, as well as further examining the ways in which these factors interact to influence performance.

Analysis 4: Analysis of Army Performance-Related Documents

The document analysis of Army-performance related literature provides important context to inform the construction and interpretation of Soldier performance statistical models.

Research Questions

Over the course of our document analysis, we examined five research questions.

- RQ1: What individual soldier characteristics contribute to unit performance?
- RQ2: Are the same or different characteristics highlighted as important across Army publications and academic team performance literature?
- RQ3: How have descriptions of performance in Army publications changed over time?
- RQ4: How well do performance concepts in items assessing Soldier psychosocial social characteristics align with performance concepts in Army publications?
- RQ5: What descriptive and prescriptive behaviors are associated with high performing Soldiers?

As previously mentioned, we concluded that the unit-level data in the PDE are unreliable below the battalion level. This analysis – specifically RQ1 – would therefore prove crucial in our investigation into how individual-level data could be leveraged to support potential models of unit performance.

Document Review

A document analysis starts with a document review. O'Leary (2014) describes the process of conducting a document review as asking questions. For example, who produced this document and when? A document review also includes:

- Exploration of each document's agenda and biases.
- Exploration of background information (e.g., document purpose)
- Exploration of content

Having a critical and contextual understanding of the documents helps to contextualize analysis and findings. Because of our varied corpus, our documents had multiple purposes and authors. The document review included general Army publications and academic literature.

Army publications have many purposes. For example:

- Army regulations are publications that sets forth missions, responsibilities, and policies; delegates authority; sets objectives; and prescribes mandated procedures to ensure uniform compliance with those policies (*ADP 1-01: Doctrine Primer*, 2019)
- Pamphlets are permanent instructional publications (ADP 1-01: Doctrine Primer, 2019). Unless mandated in an Army regulation, procedures established in a Pamphlet are for

guidance only and to establish optional or helpful methods to perform mission essential functions, define probable courses of action, and explain how something is affected.

- Pamphlets are permanent instructional publications (ADP 1-01: Doctrine Primer, 2019). Unless mandated in an Army regulation, procedures established in a Pamphlet are for guidance only and to establish optional or helpful methods to perform mission essential functions, define probable courses of action, and explain how something is affected.
- Army directives are directives or information memorandums issued by the Secretary of the Army (SECARMY) to establish or change policy or guidance for distribution and applicability across the Army (*ADP 1-01: Doctrine Primer*, 2019).
- Army directives are directives or information memorandums issued by the Secretary of the Army (SECARMY) to establish or change policy or guidance for distribution and applicability across the Army (*ADP 1-01: Doctrine Primer*, 2019).

The National Academies (2003) defines the purpose of academic publication as, generally, to communicate scientific findings and build a research community. Academic publication may also aim to influence society or the public. Often, publication improves an academic's job prospects or ability to receive promotion or tenure.

As part of the document review, we assessed whether the organization producing the documents had undergone changes or were responding to civilian or Army events using the timeline described in Part 1. The document analysis spanned almost nine decades, starting in the 1940s, so it was important to understand the events in the context of the Army documents over time. We identified the following points of scrutiny as well as possible biases in historic and current Army doctrine. We also expanded our analysis to include military culture overall where relevant.

- Historic Army recruitment tests reinforced institutional bias and maintained segregation. Due to Jim Crow laws, black recruits did not receive the same education, roles, and promotions until 1948 when President Truman issued Executive Order 9981 which mandated "equality of treatment and opportunity for all persons in the armed services without regard to race, color, religion or national origin" (General Records of the United States Government; Record Group 11, 1948). Discrimination likely existed after this period as Jim Crow laws continued in many states through the 1960s.
- The "Don't ask, don't tell" policy, which barred openly lesbian, gay, bisexual, transgender, queer, questioning, intersex, asexual (LGBTQIA+) persons from joining the military, claimed, among other things, their presence would risk unit cohesion (Feder, 2013). The policy was lifted on September 20th, 2011.
- Promotion process
    - Promotion procedures force underperforming Soldiers out (Cox, 2019)
    - The process of advocating for yourself to the Promotional Review Board may be unfair (O'Connell, n.d.)
- Gender bias
    - Gender biases in Army fitness testing used for promotions (Steinhauer, 2021)

- Army delayed promotion for two female officers until after 2020 election (Schmitt & Helene, 2021; Vindman, 2021)

Data

Document Discovery

The document analysis steps include background and preparation phases before analysis. After the document review, researchers gather texts and background information about the documents (O'Leary, 2014). We conducted a document discovery for Army performance related documents. We embarked on many methods to retrieve documents. These methods included discussion with project sponsors to identify documents of relevance of interest. We also included background or relevant materials judged as relevant to Amy performance. We also searched databases in the Army Publishing Directorate and Defense Technical Information Center, specifically the Technical Reports. Keywords associated with these database searches included "performance", "team performance", and "unit performance."

Finally, we reached out to other organizations to aid in the document discovery process. When investigating each new research question, we would search for new documents to add to the document discovery. We partnered with University of Virginia Libraries to conduct a literature sprint. The literature sprint involved a team of librarians from UVA as well as a subset of our team to find literature in a high intensity manner for one week. The literature sprint involved conducting a literature review on team science as it relates to challenges working in a virtual team science environment. This literature sprint yielded 231 materials. After screening, seven documents were added to the document inventory.

We also acquired historical Army Leadership and Leadership Development doctrine and documents from Center for the Army Profession and Leadership (CAPL). This connection yielded 29 documents from 1946 to 2019, which were added to the data inventory. CAPL also provided us with detailed metadata on the evolution of documents (e.g., introduction of new themes, changing leadership definitions) by year as tracked by CAPL. We also communicated with and received documents from the Army G-1 Publications Team and the Ike Skelton Combined Armed Research Library.

We discovered 85 documents relevant to the topics of Army performance, with 73 of these documents coming from the military sector and 13 from the civilian sector, including academic articles. See Exhibit 41 for an abbreviated example of the document discovery table. Of the military documents, the most frequently occurring types of documents were field manuals (18), technical reports (11), Army regulations (11), Army press articles (6) and pamphlets (6).

Exhibit 42.
Abbreviated Document Discovery Table

Document	Type	Source	Sector	Keywords	Description
A Big Transition: Military Service Members' Earnings and Employment After Active Duty	Literature	Goldman et al.	civilian	career, enlisted	Improving enlisted service member transitions from active duty to civilian life calls for better information about how service members fare in their transitions. The authors examined the relationship among enlisted service members' military occupations, personal characteristics, and civilian employment outcomes over the first three years after separation from active
A Survey of Army Team Operations	Technical Report	DTIC	military	team, survey, characteristics	Over the past several years, considerable progress has been made in improving the effectiveness of the soldier performing as an individual, but performance of soldiers as
ADP 6-22 (supersedes ADP 6-22 and ADRP 622, 2012 and ADRP 1, 2015) (w/ C1 2019)	Army Doctrinal Public	Center for the Army Profession and Leadership	military	leadership	
An empirical study of best practices in virtual teams	Literature - Team Science	Lurey and Raisinghani	civilian		This study explores the issue of effectiveness within virtual teams - groups of people who work together although they are often dispersed across space, time, and/or
ANALYSIS OF LEADERSHIP PERCEPTIONS AND CONTRACTING EFFICIENCY	Joint Applied Project Report	DTIC	military	develop, test, lead, indvidual	telework by the RAND Corporation, provide a foundation for the analysis of contracting efficiency metrics and leadership perceptions
AR 600-100 (supersedes AR 600-100, 1986)	Army Regulation	Army Profession and Leadership	military	leadership	

To aid in document screening and selection, we tracked a variety of metadata for each document. This metadata included:

- Document Name
  - Including Army Doctrine Type and Series Number
- The Type of Document
  - Army doctrine inherited types from Army publications
  - For non-Army publications, we developed a categorization typology
- The Source of the Document
  - e.g. Army Publishing Directorate, DTIC
- The Sector
  - Military or Civilian
- Keywords associated with the document
- Description
  - Often the executive summary or abstract
- # of pages
- Document date
  - Note: The document date was not initially tracked, but as our project evolved, we became more interested in the evolution of Army doctrine over time. The document discovery, then, is inconsistent in its treatment of documents which have been updated year after year as multiple documents.

The Document Discovery Table does not include all sources that were used for analysis. Namely, item text from the Global Assessment Tool (GAT) and Tailored Adaptive Personality Assessment System (TAPAS) are not included in the Table. Because these are not documents per say, it is difficult to tag them using the prescribed structure. Furthermore, these items are already inventoried in detail in the Conceptual and Methodological Profiling Document.

To properly compare survey item text to the rest of the corpus, we need syntactic alignment items syntactically. To achieve this, we constructed statements composed of the given survey item stem and text and a response. For example, the item stem, "Here are a number of words that describe different feelings and emotions. How often you have felt this way during the past four weeks?" and the item text, "Joyful" could be combined with the response, "Most of the time" to produce the statement, "I have felt joyful most of the time during the past four weeks." Using this procedure, we created statements reflecting both productive and counterproductive performance (e.g. combining the previous stem and text with the response "Never" would create the counterproductive performance statement, "I have never felt joyful during the past four weeks.")

Document Screening

The document discovery table was revisited and augmented as we went through each research question. Documents were screened and selected based on the relevance to the given research question and with analytical limitations in mind. For example, documents that contained relevant material within charts or infographics were not selected for quantitative document analysis, as key features would be lost in the text extraction process.

Methods

Qualitative Document Analysis

Document analysis is a form of qualitative research in which documents are interpreted by the researcher to give voice and meaning around an assessment topic (Bowen, 2009). Analyzing documents incorporates coding content into themes similar to how focus group or interview transcripts are analyzed (Bowen, 2009). Public records, such as Army publications, are one of the primary types of documents identified by O'Leary (2014), which can give us insight into organizations. Thematic analysis, the type of analysis chosen for this study, is akin to pattern recognition. When performing thematic analysis researchers code documents for emergent themes to create categories (Bowen, 2009). Document analysis usually includes triangulation, which is an analysis of convergent or divergent themes within a different corpus or media (Bowen, 2009).

A team of four researchers were the primary individuals performing qualitative analysis. We chose to implement a hybrid inductive-deductive approach to identify themes. Inductive analysis starts by examining the data. Deductive analysis organizes data based on pre-existing ideas and research. Hybrid analysis uses a combination of both approaches.

To identify themes, we developed a document coding scheme to identify patterns in the documents that point to themes. Before we began qualitative analysis, we identified a subset of relevant codes based on background knowledge and sponsor input (a-priori themes). We selected 10 documents from the corpus that would be relevant to our given research questions (RQ1 and RQ2). Each researcher was assigned two or three documents to read initially. The initial read-through of these documents was aimed at identifying emergent codes. After our initial reading, we met as a group to refine and consolidate codes and develop a final list of codes (see Exhibit 42). After developing our final list of codes, each document was thoroughly coded again to align with the reconciled coding scheme.

Exhibit 43.
Final Qualitative Analysis Codes

Code	Type	Definition	Example
Qualities of a Leader	A Priori	Specifies a characteristic that is important to being a good leader. Clearly connects this quality to success in leadership. Code is always paired with specific qualities.	Leaders can display character by completing tasks on time and modeling standards for performance, personal appearance, physical fitness, and ethics. They can also display character by modeling sound judgement and reasoning, determination, persistence, and patience
Leader development	A Priori	Deliberate, continuous, and progressive process, founded in Army Values, which grows others into competent, committed, professional leaders of character.	Leader development is a continuous lifelong learning process beginning with pre-appointment training and education.
Organizational Trust	A Priori	Assesses three dimensions of organizational trust: ability, benevolence, and integrity. Encapsulates trust in the broader organization of the Army, not in a particular individual. Other terms include culture of trust includes indoctrination.	We all have an important job to do, necessary to the unit's missions. Soldiers throughout the Army perform the duties of medics, infantrymen, cooks, truck drivers, mechanics, legal clerks, aviators, and other occupations. We bring fuel to the tanks, scout for the enemy, listen to the enemy's signals, and close in and defeat the enemy. We defend against air attacks, ensure Soldiers are paid accordingly, and process awards to recognize other Soldier's accomplishments. As important as individual Soldier tasks are, we all know that these efforts support a team and that the whole is greater than the sum of its parts. Every team has a leader, and that leader is responsible for what the team does or fails to do. That is why obeying orders is the critical essence of discipline.

Code	Type	Definition	Example
Diversity in Team	Emergent	Differences in team members in areas such as background, religion, race, socioeconomic status, or areas of expertise. Includes diversity in skill and background.	Variation in upbringing, culture, religious belief, and tradition is reflected among those who choose to serve in the Army. Such diversity provides many benefits for a force globally engaged around the world. Good leaders value this diversity of outlook and experience and must treat all individuals with the inherent dignity and respect due every person.
Negative Qualities	Emergent	If paired with a code, it talks about that other code in a bad way. If on its own, it describes a negative quality or behavior like over drinking.	Misunderstands or fails to perceive nonverbal cues. Ideas not well organized or easily understandable. Speaks without considering listener interest. Information dissemination is inconsistent or untimely.
Team Succes	A Priori	Describe what it means to be a successful team or how team success may be measured.	High performance teams demonstrate mental agility (see ADRP 6-22) in their willingness to approach problems from different viewpoints and to hold and work on opposing ideas until identifying the best solution
Shared Vision	Emergent	Common and clear goals and vision that members of a team identify with and uphold. This is on a smaller level, for example your unit knows exactly what your expected outcome is.	Shared vision exists when members of the team have a common understanding of the following: The overall mission or objective. Goals and sub-goals of the mission. Strategies for reaching those goals. Team members' strengths and weaknesses. Values and preferences of the team as a whole and among the individual members. The roles each member will play and want. The big picture (such as how this particular mission contributes to a larger purpose).
Team Building	A Priori	Actions or processes that encourage members to work together.	The team-building approach addresses the interaction and relationships among team members

Code	Type	Definition	Example
			essential to collaboration and critical to high performance. It accelerates collaboration by allowing the team members to further develop qualities of shared vision, trust, competence, and confidence among team members. It is a streamlined approach designed to break through barriers and boundaries that stagnate team development.
Responsibility	A Priori	Individuals are dependable, reliable, and make every effort to keep their promises.	The general responsibilities of a Soldier are as follows: Obey the lawful orders of NCOs and Officers. Treat others with dignity and respect. Complete each task to the very best degree possible and not just to standard. Maintain a military appearance. Maintain individual physical fitness standards and readiness. Maintain individual equipment and clothing to standard.
Cooperation	Emergent	Measures agreeableness, trust, skepticism and suspicion; the extent an individual is easy or difficult to get along with. Includes cohesion.	Exemplifies a positive attitude and expectations for a productive work environment. Conveys a priority for development within the organization. Encourages innovative, critical, and creative thought. Leverages lessons learned to improve organization.
Influence	Emergent	Having an effect or influence on another person or organization.	Projects self-confidence and inspires confidence in others. Models composure, an outward calm, and control over emotions in adverse situations. Manages personal stress and remains supportive of stress in others.
Intellectual Efficiency	A Priori	Measures a person's ability to analyze and process information, astuteness or obtuseness. For innovation,	Demonstrates expert-level proficiency with technical aspects of job. Demonstrates understanding of joint, cultural and geopolitical knowledge. Conveys knowledge of

Code	Type	Definition	Example
		use this & adaptability Includes efficiency in skill.	technical, technological, and tactical systems to subordinates and others.
Negation of Positive Qualities	Emergent	This makes any positive quality negative. It must also be selected with a positive quality. For example, if a passage described poor communication, you would select this and then communication.	Communication frequency was negatively related to both ROI and sales growth, contrary to what had been hypothesized. Smith et al. (1994) suggested that perhaps communication frequency is indicative of high levels of conflict in the group, and that the time spent on group maintenance detracts from performance.
Adaptability	A Priori	Assesses the ability to alter one's course, and perceived cognitive flexibility. For innovation, use this & intellectual efficiency.	Models a flexible mindset and anticipates changing conditions. Engages in multiple approaches when assessing, conceptualizing, and evaluating a course of action.
Team Orientation	A Priori	Measures a person's tendency to prefer working in teams and make people work together better.	The success of the U.S. Army is related directly to the quality of the professional relationships between its Soldiers, NCOs, and officers. The relationship between an individual Soldier, team, squad, and platoon with their NCO/officer is the cornerstone of our Army. When that bond forms, it can have the single most important impact on unit effectiveness and efficiency. Conversely, if that bond is broken, it can have a devastating impact on morale, esprit de corps, readiness, and mission accomplishment.
Commitment to Serve	A Priori	Measures a person's level of identification with the military and their strength of desire to serve their country.	The pride, esprit, and ethos required of Soldiers as members of the Profession of Arms may require them to sacrifice themselves willingly to preserve the Nation, accomplish the mission, or protect the lives of fellow Soldiers.

Code	Type	Definition	Example
Communication	A Priori	Exchange of information between individuals. Includes verbal and non-verbal.	The questioning creates dialogue, which improves awareness and leads to a higher degree of actionable understanding. Through the dialogue, the team develops shared understanding and trust. When coaching a team exercise: Frame initial questions to clarify the task and the situation. There will often be disagreements on both. Find common ground. Try to get the team to come to agreement on one issue. Build to another agreement. When there is disagreement, try to get at the specifics of what is disagreed upon, before building to said Agreement.
Integrity	Emergent	Having honest and strong moral principles.	Character is comprised of a person's moral and ethical qualities, helps determine what is right, and gives a leader motivation to do what is appropriate, regardless of the circumstances or consequences. It determines who people are, how they act, helps determine right from wrong, and choose what is right.

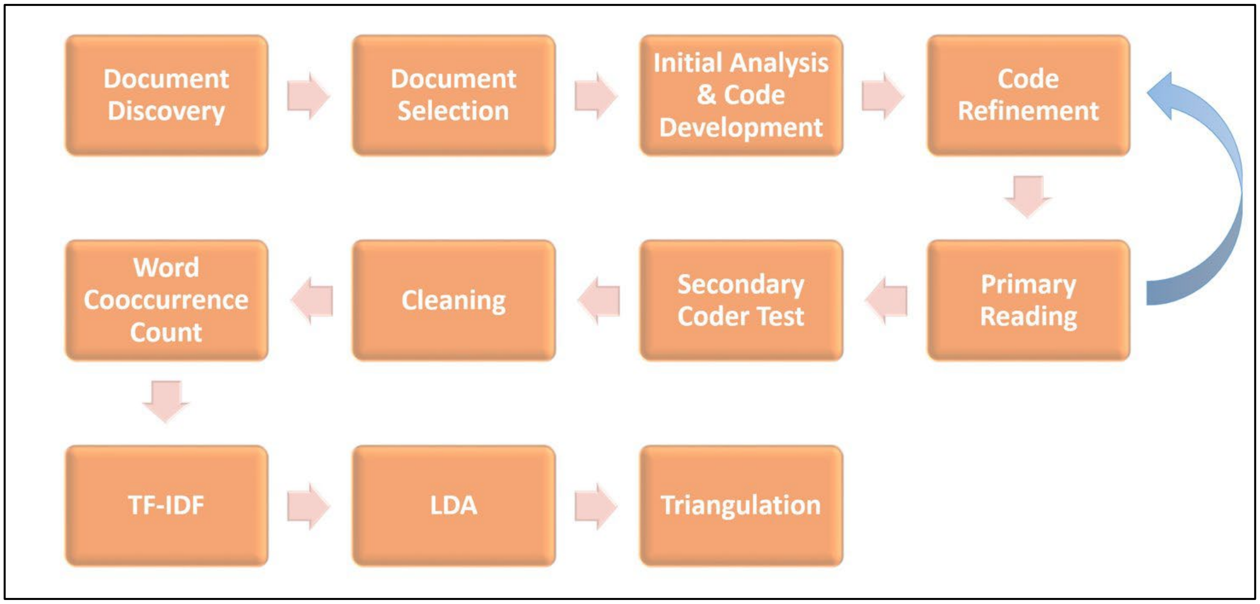
Note: Codes are categorized by type (emergent or a priori). For each code a definition and an example are provided.

After coding each document, we tested for inter-rater reliability. A group of five researchers external to the project were asked to code a subset of the documents to test code agreement. The coding and testing of documents were performed in Dedoose, an application for mixed-methods research. Two tests were in the Dedoose Training Center created to analyze the 18 codes in Exhibit 43. The five secondary coders each performed both tests. Results were analyzed by code and evaluated based on Dedoose's interpretation of Cohen's Kappa to evaluate inter-rater reliability.

Triangulation

To triangulate on the findings of our qualitative analysis, we performed topic modeling on the cor. The topic modeling aimed to see if codes and themes identified from manual analysis converged or diverged with results derived from Natural Language Processing (NLP) (RQ1 and RQ2). Exhibit 43 shows the process for triangulation for RQ1 and RQ2. Miner et al. (2023) perform a similar analysis to compare the results of a previous qualitative analysis on a corpus of written experiences of workplace discrimination to an automated, topic modeling approach to analyzing the data. They find mixed results regarding the convergence of qualitative and quantitative findings. However, they point out the strengths that NLP has for analyzing large corpora in a more efficient manner than manual qualitative coding. Topic modeling using Latent Dirichlet allocation (LDA) is further explained in the quantitative document analysis methods section (Blei et al., 2003). Based on results from the LDA model, manual codes were matched with the topics to identify convergent and divergent findings.

Exhibit 44.
Document Analysis and Triangulation Process



Note: This figure shows process of triangulation for analyses involving document analysis and NLP. After performing document analysis steps, documents were cleaned and descriptive NLP was performed, leading to triangulation of themes.

Quantitative Document Analysis

Data Processing and Cleaning

The corpus was initially uploaded to R Studio as images (pdf) and converted to text (txt) files using the pdftools R software package. The pdftools package uses a combination of text parsing and optical character recognition to convert documents to text. After converting the documents to text, we began the process of cleaning the text for analysis.

The following data cleaning steps account for all possible cleaning procedures for analysis. Though, it should be noted that not all cleaning steps are necessary or helpful for each analysis. Where appropriate, the cleaning steps omitted from a particular analysis are clarified.

- Text is converted to all lowercase characters
- Non-alphabetic characters are removed
- Excess whitespace is stripped

- Removal of stop words using the stringr stop word dictionary (e.g., the, and, of)
    - Lemmatization (also called stemming) of words (e.g. running and runs become run)
    - At this point, “words” are often referred to as tokens, as the stems of words may not be words themselves. While token can mean any component parts of text, in this report tokens exclusively refer to the text cleaned after this stage.
- Removal of last names and personally identifiable information (PII) using the Scrubadub software package
- Removal of tokens with fewer than four letters. Tokenization splits the raw text into small chunks of words or sentences, called tokens (Gunjal, 2020).

In some cases, specific subsets of documents were extracted for analysis. The documents were subset either by using regular expressions or copying and pasting relevant sections for analysis.

To prepare text for analysis, often the data are shaped into specific formats. One of these formats is a document-term matrix. A document-term matrix tracks the presence of tokens within certain documents or paragraphs. A document-term matrix, then, captures relationships between words without the context in which they appear. Not all analyses required a document-term matrix.

Term Frequency

Term frequency describes the number of times a term (token) is used within a given paragraph or document. Term frequency, using a document-term matrix, yields a high-level description of the concepts described in a paragraph or document. Term frequency can be thought of as a crude representation of what concepts may or may not be important, and it is assumed that more frequently occurring words bear more weight on the document and less frequently occurring words bear less weight.

In our analysis, we analyzed the frequency of several different types of terms. Often term frequency was the first descriptive step to analyze documents. The term frequency analyses that are described in detail in this report are the pronoun frequency and the verb frequency.

We analyzed pronoun frequency to understand how performance documents over time changed the unit of analysis they were referring to (RQ3). Most importantly, this included tracking changes in the frequency of individual versus collective pronouns (I/me/mine versus we/ours/us). We also tracked individual gendered pronoun frequencies (she/her/hers and he/him/his) (Gaucher et al., 2011).

We also analyzed the frequency of verbs within specific task performance-related documents (RQ 5). To understand the breakdown of different performance behaviors, we extracted the verbs from these texts and analyzed their frequency. We used the udpipe software package to tag parts of speech, then subset the verbs from these documents. Udpipe is developed by researchers within the Institute of Formal and Applied Linguistics at Charles University (Straka & Straková, 2017).

Co-occurrence

Co-occurrence is an extension of term frequency and also uses a document-term matrix. The goal of co-occurrence is to introduce more nuance into the presence of terms themselves by searching for terms that frequently co-occur in a sentence, paragraph, or document. Note this is different than analyzing bi-grams, which are two words that occur directly in sequence. Co-occurrence captures two tokens that appear together in the same sentence, paragraph, or document regardless of sequential ordering. Co-occurrence then can tell us crudely what concepts might be related to each together in a given text or give more context to frequently occurring terms.

Co-occurrence in this analysis was used to answer RQ1. For the corpus of Army-performance related documents, co-occurrence was performed to understand high-level trends in the co-occurrence of tokens or ideas. This yielded a high-level overview of the concepts discussed within documents are how the strength of their connection (by frequency).

Co-occurrence was also used to investigate performance behaviors (RQ 5). After analyzing the frequency of verbs alone, we analyzed the co-occurrence of verbs with other tokens within task items. These yielded verb phrases or ideas that frequently occurred in pairs within task items.

Term Frequency-Inverse Document Frequency

Term Frequency-Inverse Document Frequency (TF-IDF) is also an extension of term frequency. TF-IDF is calculated using a document-term matrix by multiplying the number of times each term appears in each document and the inverse document frequency of the term in the corpus of documents. TF-IDF output, then, shows which terms make each document unique from other documents in the corpus.

Given the strengths of the method, we used TF-IDF to answer RQs 2 and 3, which seek to tease out differences or changes in documents. TF-IDF was performed for an initial corpus of Army documents and team science literature to understand the overall themes in performance and how each document was differentiated. Later, TF-IDF was used in the corpus of PAM 600-3 documents over time to understand changes or differentiations in the terms used to describe Officer promotion over time.

Topic Modeling

Topic modeling is a machine learning technique designed to find topics within a corpus of text. The topic modeling method used in this analysis was Latent Dirichlet Allocation (LDA) (Blei et al., 2003), one of the two models tested in our earlier analysis examining performance scales (see Part 1: Analyzing performance items with natural language processing (NLP)). LDA is a form of topic modeling that is used to find topics within a corpus of documents and identify the words associated with each topic. LDA is a supervised machine learning technique. The number of topics to search for is a parameter set by the researcher.

We used LDA to answer RQ1. Specifically, we performed this analysis to triangulate on the themes identified from the manual document analysis. For the purpose of this project, we selected 18 topics identified when manually reading the documents.

Sentiment Analysis

Sentiment analysis is a text analysis technique designed to extract the emotional tone of a text (Isasi, 2023). Sentiment analysis may be able to pick up on underlying patterns in the sentiment of a document based on word choice. Sentiment dictionaries are built by scholars and developers so that large-scale sentiment analysis can be efficiently executed. To perform sentiment analysis, we used the R package syuhet and the syuhet sentiment dictionary. Using this software package, we tagged tokens with two sentiments (positive and negative) and emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, and trust). This sentiment dictionary was imported from the package. Term frequency methods are then used to understand what the percentage of tokens in each document corresponded to the given sentiments and emotions.

We performed sentiment analysis to answer RQ3 to understand how the ideas around performance and the way it was being discussed changed over time by sentiment.

Large Language Models

The final analysis method used was large language models. Large language models are deep learning models that are trained on large amounts of text (Google, 2023). In our case, we used pre-trained large language models to understand the contextual relationships between text in our corpus. Large language models can be used to contextualize relationships between words in a sentence and across documents and identify emergent and constructs that are hard to observe through manual qualitative analysis.

Large language models capture the contextual relationships between words in a sentence and across documents, which we use to identify emergent and unobservable constructs across disparate data sources. We link these latent concepts to specific prescriptive and descriptive performance behaviors, as well as the characteristics Soldiers are expected to manifest through their behavior (e.g., loyalty, bravery, leadership). Incorporating historical records allows us to track changes in the description of the concepts and behaviors across time.

The output of a large language model tokenization is a vector, or a numerical representation of the document. These vectors can be analyzed using different techniques. For our analysis, we used a distance measure (cosine similarity) and clustering using K-nearest neighbors to understand relative differences between documents.

The first large language model used in this analysis is Bi-directional Encoding using Transformers, or BERT. BERT was developed by Google and is available open-source on Hugging Face. We used the most common version of BERT, bert-based-uncased for our analysis. Since this language model is bidirectionally trained, it can have a deeper sense of language context and flow than single-direction. BERT is designed to help computers understand the meaning of ambiguous language in text by using surrounding text to establish context.

Generative Pre-trained Transformers (GPT) is a series of large language models developed by OpenAI, who is well-known for having created the GPT series as well as the online accessible version, ChatGPT. GPT-2 is more of a black box than BERT, as later versions of the model are not publicly accessible or well documented. GPT-2, then, is the most recent version of the model that is publicly accessible on HuggingFace. It is a transformer-based language model designed to understand and generate human-like text by predicting the likelihood of words or sentences given the context provided to it. GPT-2 can perform various language-related tasks, such as generating coherent text, answering questions, completing sentences, and more. Its ability to generate contextually relevant responses has made it a powerful tool for NLP applications.

Results and Discussion

The results presented correspond with each research question.

RQ1: What characteristics of an individual Soldier contribute to unit performance?

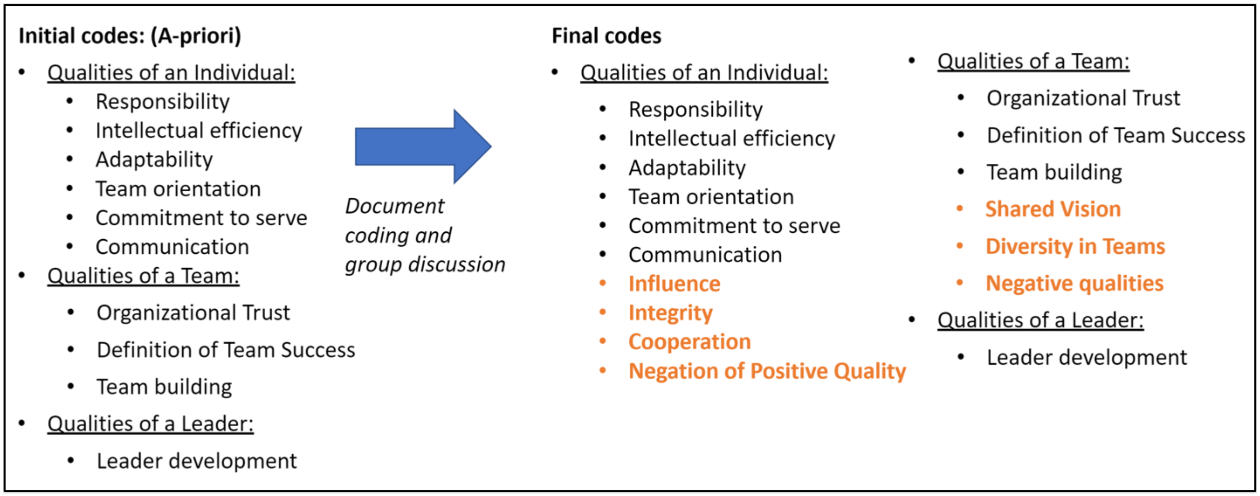
A1. Selected Results from Qualitative Document Analysis

The first analysis performed to study the characteristics of individual Soldiers that contribute to unit performance was qualitative document analysis. The corpus of documents in this case (Corpus 1) was a blend of Army doctrine and literature from the team science field. Corpus 1 includes:

- Center for Army Lessons Learned, 2021
- Cohen & Bailey, 1997
- Svyantek et al., 1999
- U.S. Department of the Army (1946, 2014a, 2015d, 2015b, 2015e, 2019d, 2019c, 2019a)

Codes. During our document analysis process, our coding scheme evolved, as is common in qualitative analysis. At the outset, the team identified 10 a-priori codes for qualitative document analysis. A-priori codes were determined based on prior knowledge about Army individual and unit performance, as well as standardization with items from the Global Assessment Tool (GAT), or characteristics that we already know the Army measures and could be used to proxy performance. At the end of code reconciliation and testing, we identified 18 codes for qualitative document analysis. Exhibit 44 shows the evolution of the a-priori codes into the final codes, with orange representing the emergent codes in the final set.

Exhibit 45.
The Process of Qualitative Code Refinement



Note: This figure shows the process of refining initial codes to final codes for qualitative document analysis. Orange codes emerged out of document coding and group discussion.

Through our document coding procedure, we identified 320 excerpts. These excerpts were tagged with 506 applications of the 18 codes.

Dedoose reports inter-rated reliability of code testing using the Cohen's Kappa statistic for each tester individually and as an average for all testers. We focused our analysis on the average Cohen's Kappa to analysis agreement by code wholistically instead of focusing on individual raters. Exhibit 45 shows the results of each test individually and the average Cohen's Kappa between the two tests.

Exhibit 46.
Results of Qualitative Coding Tests

Code	Average Cohen's Kappa	Test 1 Cohen's Kappa	Test 2 Cohens' Kappa
Diversity in Team	0.482	0.492	0.472
Team Orientation	0.132	0.154	0.11
Responsibility	0.299	0.328	0.27
Commitment to Serve	0.482	0.508	0.456
Communication	0.511	0.462	0.56
Shared Vision	0.108	0.128	0.088
Influence	0.431	0.338	0.524
Negation of Positive Qualities	0.325	0.248	0.402
Adaptability	0.418	0.628	0.208
Intellectual Efficiency	0.126	0.348	-0.096
Team Building	0.143	-0.014	0.3
Cooperation	0.146	0.342	-0.05
Integrity	0.288	0.456	0.12
Definition of Team Success	0.177	0.306	0.048
Leader Development	0.167	0.072	0.262
Organizational Trust	0.127	0.242	0.012
Negative Qualities	0.183	0.132	0.234
Qualities of a Leader	0.286	0.152	0.42

We can interpret Cohen's Kappa using the following ranges:

- ≤ 0 No agreement
- 0.21–0.40 Fair agreement
- 0.41– 0.60 Moderate agreement
- 0.61–0.80 Substantial agreement
- 0.81–1.00 Almost perfect agreement

We found that Cohen's Kappa for most codes spanned more than one interpretation range. Across two tests for our code applications, our agreement was categorized as the following:

- One code scored fair to substantial agreement
    - Adaptability
- Three codes consistently scored moderate agreement
    - Diversity in Team, Commitment to Serve, and Communication
- Two codes scored fair to moderate agreement
    - Influence, Negation of Positive Qualities
- One code scored slight to moderate agreement
    - Qualities of a Leader
- One code scored none to moderate agreement
    - Integrity

- Two codes scored consistent fair agreement
    - Team Orientation, Responsibility
- Four codes scored slight to fair agreement
    - Definition of Team Success, Leader Development, Organizational Trust, Negative Qualities
- One code scored consistent slight agreement
    - Shared Vision
- Three codes scored none to fair agreement
    - Intellectual Efficiency, Team Building, Cooperation

After analyzing the test results, we discussed confusion in the coding scheme as a team. We refined codes and excerpts where applicable, arriving at the final coding scheme and coded excerpts.

Themes. Below, we highlight noteworthy findings from our analysis with a deeper examination of several converging and diverging themes from across multiple documents.

Negative Qualities vs. Negation of Positive Qualities. One emergent code was Negation of Positive Qualities. Both negative qualities and negation of positive qualities capture counterproductive performance. Exhibit 46 shows a comparison of two excepts tagged with these codes. On the left, the excerpt highlights a situation in which too much communication was associated with negative team outcomes (i.e., communication was seen as a 'negative quality' for a group to possess). The excerpt on the right highlights the opposite case, instances where too little (or a lack of) communication and/or poor-quality communication is associated with negative team performance. Taken together, we see that communication is a crucial element in determining team success, but that there is an important contextual element to understanding exactly how communication affects success. For example, if someone is a poor communicator and a team is suffering for it, simply instructing the individual to communicate more is unlikely to resolve the issue. In fact, it could make the situation worse.

Exhibit 47.
Comparing Negative Qualities and Negation of Positive Qualities Excerpts

Communication. Communication frequency was negatively related to both ROI and sales growth, contrary to what had been hypothesized. Smith et al. (1994) suggested that perhaps communication frequency is indicative of high levels of conflict in the group, and that the time spent on group maintenance detracts from performance.

COMUNICATES
Misunderstands or fails to perceive nonverbal cues. Ideas not well organized or easily understandable. Speaks without considering listener interest. Information dissemination is inconsistent or untimely.

Note: This figure shows two excerpts treat the counterproductive quality of communication differently. The left excerpt refers to negation of a positive quality, while the right excerpt refers to a negative quality.

Adaptability vs. Intellectual Efficiency. Exhibit 47 shows a comparison between excepts coded as Adaptability and Intellectual Efficiency from Army publications. The excerpts on the left were coded as 'Adaptability' while the excerpts on the right were coded as 'Intellectual Efficiency.' Distinguishing between these two codes proved difficult during the qualitative analysis because, as is highlighted in the underlined segments, the Army defines both qualities using similar language. We were hesitant to combine these codes, though, because we know from the GAT that the Army sees these two characteristics as conceptually different.

Ultimately, we refined our coding scheme to distinguish between the two through the use of language such as that highlighted in yellow. We determined that to be 'adaptable' one must be able to think critically, while 'intellectual efficiency' rests on a person's ability to analyze and process information to include efficiency in skill. Efficiency in skill includes experiences, training, education, and so on (as highlighted in the top right excerpt), which are not necessarily required in order to be adaptable. Intellectual efficiency includes reasoning and brain power, while adaptability making quick decisions and flexibility thinking, which is crucial for a Soldier to thrive in a 'high operational tempo.'

Exhibit 48.
Comparing Adaptability and Intellectual Efficiency Excerpts

"Research on transactive memory systems, or knowledge of the expertise each team member has, has found that team's function better when they know who is good at doing what. For example, engineering teams who train together tend to perform better because members are better able to specialize their knowledge to the benefit of the team, and are able to use each other as knowledge resources. The influence of strong transactive memory systems on team performance is broad. "
- The Importance of Individual Member to Team Success (Non-Army)

"To operate at high performance levels, organizations are challenged when building teams whose members cross multiple boundaries. Team building seems intuitive when cultures, values, and language are shared; however, in some teams, diverse backgrounds, agendas, perspectives, and situational perspectives complicate things and can even render a team dysfunctional."
- Leaders Guide to Team Building (Army)

"Again, these numbers suggest that owners and managers of some of the most specialized teams in the world operate under the implicit assumption that individual employees impact team performance, not just because they are part of a high performing team, but because of their unique individual characteristics"
- The Importance of Individual Member to Team Success (non-Army)

"Variation in upbringing, culture, religious belief, and tradition is reflected among those who choose to serve in the Army. Such diversity provides many benefits for a force globally engaged around the world. Good leaders value this diversity of outlook and experience and must treat all individuals with the inherent dignity and respect due every person. "
- Army Leadership and the Profession (Army)

Note: This figure shows several excerpts that exemplified the difficulties encountered in discerning the differences between adaptability and intellectual efficiency in the documents.

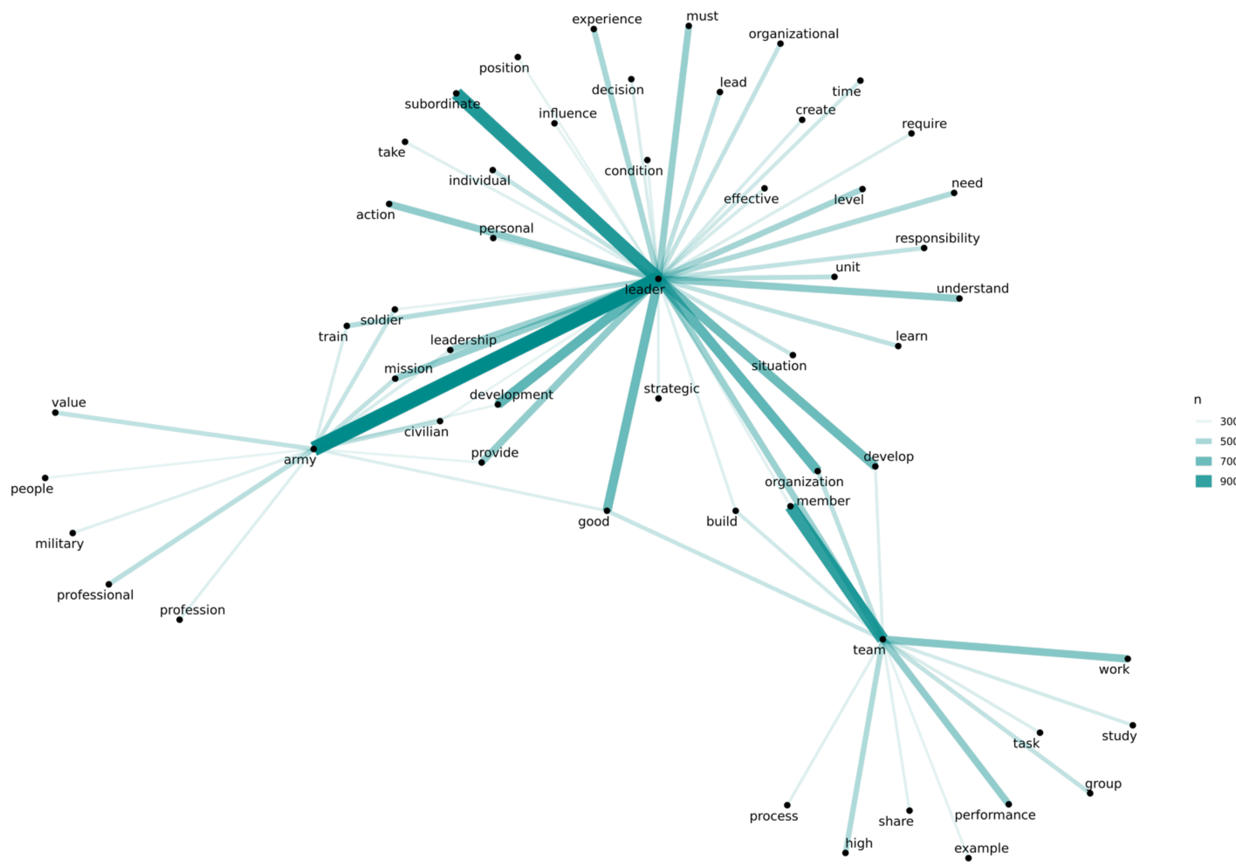
A2. Co-occurrence

The following analyses constitute our triangulation on the findings from the qualitative document analysis using a co-occurrence analysis. Corpus 1 was also used for A2.

Exhibit 48 shows the results of the co-occurrence analysis. Some of the largest co-occurrences were between Army and leader, leader and subordinate, and team and member. Words that appear together often may be conceptually correlated.

The co-occurrence findings did not necessarily support the findings from the qualitative document analysis. The overall co-occurrence patterns are not detailed enough to identify the individual characteristics that contribute to unit performance. Rather, we interpreted the results of the co-occurrence as a signal that our corpus is well selected to study this topic. If anything, the co-occurrence underscores that leadership themes are highly connected to performance within this corpus.

Exhibit 49.
Term Cooccurrence



Note: Term cooccurrence is shown within the A1 corpus. A darker and thicker line indicates terms with a higher cooccurrence (higher 'n' or number of correlations). For the purposes of this study, it is important to note that the colors used for graphs were selected arbitrarily and have no correlation to any derived meaning.

A3. Latent Dirichlet Allocation (LDA)

A3 builds on our triangulation of the findings from the qualitative document analysis using topic modeling. Corpus 1 was also used for A3.

Because we arrived at 18 codes within our qualitative coding scheme, we ran an 18-topic LDA topic model. We did not assume that the topic modeling results would match the qualitative coding exactly, but we did find certain topics aligned with the themes identified in the document analysis. Exhibit 49 shows the results of the LDA topic model. We discerned a label for each topic using the high beta value terms as well as the codes from the qualitative analysis. Beta values indicate the number of words in a topic. The higher the beta, the larger the number of words in the corpus that are in the topic. Many of the topics strongly reflected our qualitative analysis codes, providing convergent evidence to the veracity of the qualitative analysis. Some of the codes – such as

Qualities of a Leader, Shared Vision, and Team Success – were associated with multiple topics. This may suggest that these qualities are particularly important to successful team performance in the Army. It could also suggest that machine analyses determine these qualities to be more nuanced than our manual analysis. The final interpretation is that these topics are not actually conceptually distinct, suggesting that 18 topics may be too many for this corpus.

Exhibit 50.

Results of the 18 Term Topic Model

Note: Topics are labeled based on researcher discretion. For the purposes of this study, it is important to note that the colors used for graphs were selected arbitrarily and have no correlation to any derived meaning.

It is important to note that we manually assigned the names of the topics associated with the LDA output. For this reason, there is some subjectivity involved in identifying the exact alignment between codes and topics. Overall, there was convergence between qualities identified in the manual qualitative analysis and the topic modeling:

- 14 manual codes appeared in the topics

- Qualities of a Leader, Leader Development, Negative Qualities, Team Success, Shared Vision, Team Building, Responsibility, Influence, Intellectual Efficiency, Adaptability, Team Orientation, Commitment to Serve, Communication, Integrity
- Four manual codes did not appear in the topics
  - Organizational Trust, Diversity in Team, Cooperation, Negation of Positive Qualities
- One topic did not appear in manual codes
  - Family support

RQ2: Are the same or different characteristics highlighted as important across Army publications and academic team performance literature?

A4. Selected Results from Qualitative Document Analysis

Our qualitative analysis was also leveraged to answer RQ2. In this analysis, codes were compared to discover differences and similarities in characteristics highlighted as contributing to team performance Army publications versus the academic literature. This analysis also used Corpus 1.

A key difference we noticed was how Army publications discussed how diversity contributes to unit performance versus academic literature. Exhibit 50 shows a comparison of excepts coded as diversity. In the excerpt on the left, which is from the team performance academic literature, we diversity discussed in the context of individuals' knowledge, functions, and skills. Per the academic literature, having a diverse team contributes positively to team performance. In the Army publications excerpts on the right, diversity is discussed in the context of individuals' background, upbringing, and culture. It is critical in the Army context for diversity to be handled with great leadership in order for teams to benefit and avoid dysfunction.

Exhibit 51.

Comparing Excerpts on Diversity in Army and Civilian Documents.

"Research on transactive memory systems, or knowledge of the expertise each team member has, has found that team's function better when they know who is good at doing what. For example, engineering teams who train together tend to perform better because members are better able to specialize their knowledge to the benefit of the team, and are able to use each other as knowledge resources. The influence of strong transactive memory systems on team performance is broad. "
- The Importance of Individual Member to Team Success (Non-Army)

"Again, these numbers suggest that owners and managers of some of the most specialized teams in the world operate under the implicit assumption that individual employees impact team performance, not just because they are part of a high performing team, but because of their unique individual characteristics"
- The Importance of Individual Member to Team Success (non-Army)

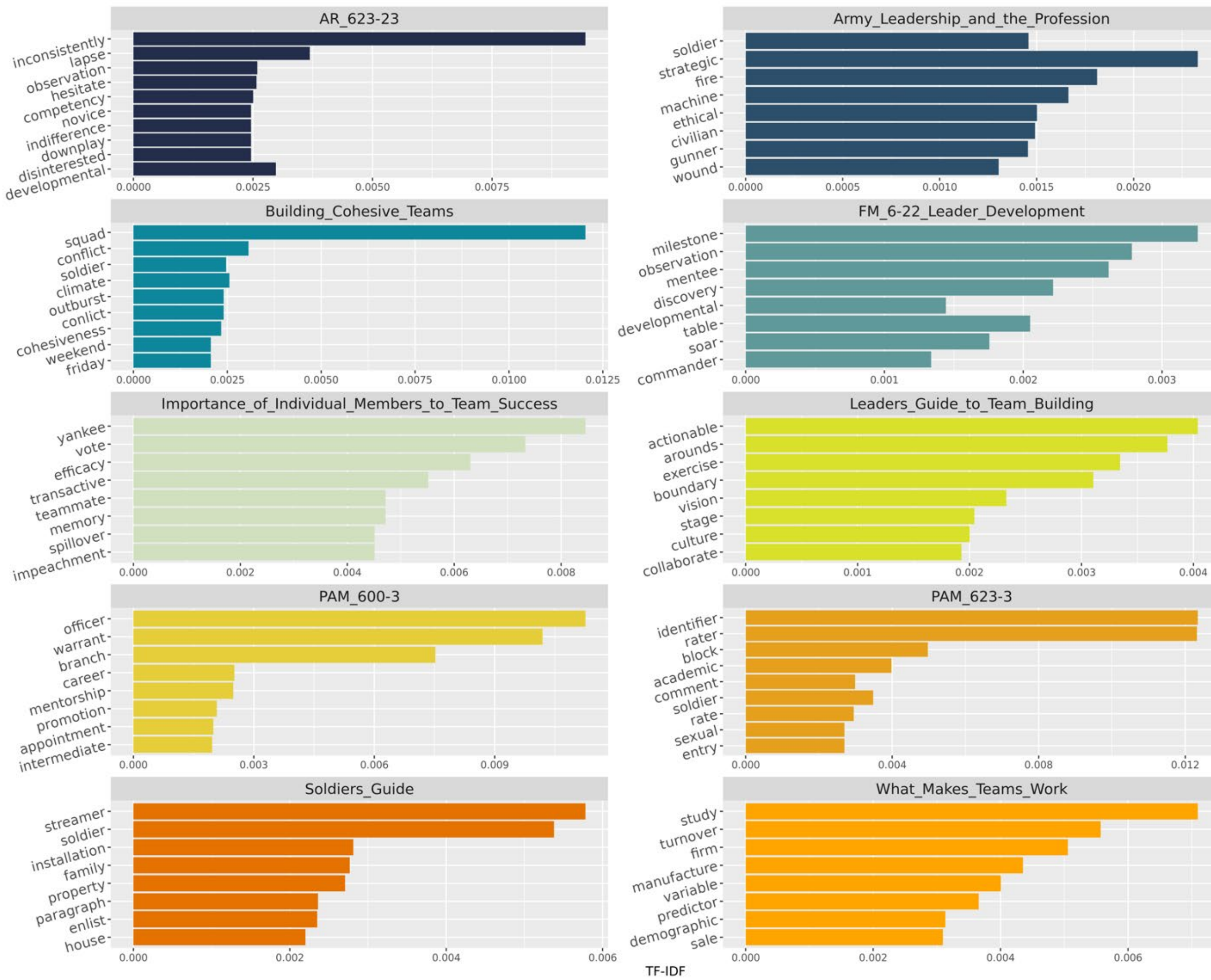
"To operate at high performance levels, organizations are challenged when building teams whose members cross multiple boundaries. Team building seems intuitive when cultures, values, and language are shared; however, in some teams, diverse backgrounds, agendas, perspectives, and situational perspectives complicate things and can even render a team dysfunctional."
- Leaders Guide to Team Building (Army)

"Variation in upbringing, culture, religious belief, and tradition is reflected among those who choose to serve in the Army. Such diversity provides many benefits for a force globally engaged around the world. Good leaders value this diversity of outlook and experience and must treat all individuals with the inherent dignity and respect due every person. "
- Army Leadership and the Profession (Army)

A5. Term Frequency-Inverse Document Frequency (TF-IDF)

To further understand differences in Army publications and academic literature, also we employed a TF-IDF approach (Griffiths & Steyvers, 2004). This analysis also used Corpus 1. As previously highlighted, TF-IDF shows document-specific terms that are relatively distinct in comparison to a corpus as a whole. A higher TF-IDF indicates a more distinct term. Exhibit 51 shows the results of this analysis.

Exhibit 52.
Term Frequency-Inverse Document Frequency Results for the 10 Document Corpus.



Note: For the purposes of this study, it is important to note that the colors used for graphs were selected arbitrarily and have no correlation to any derived meaning.

We found terms distinct to Army publications were mostly Army-specific terms, such as warrant, squad, and Soldier. The team performance academic literature included some civilian specific terms, like firm. Interestingly, terms like demographic and culture, which align with diversity concepts, were identified as distinct to the academic team performance literature. We also find that AR 623-23 seems to include several distinct terms relating to counterproductive performance.

RQ3: How have descriptions of performance in Army publications changed over time?

As background preparation for this analysis, we performed a manual review of the changes in the PAM 600-3 over a 20-year period. In the last twenty years, PAM 600-3: Commissioned Officer Professional Development and Career Management has undergone changes in name and content, although the theme of the document, to guide officers so they may excel in their careers, has remained. In 2005, the title changed to its current form from Commissioned Officer Development and Career Management. New functional areas (e.g., Psychological Operations and Electronic Warfare) have been added as Army needs required and education training options have also changed as courses have been replaced and new leadership paths have emerged.

However, the focus on the importance of educational opportunities to train leaders across functional areas has remained. In fact, in 2010 the chapter content was standardized across branches and functional areas, which may emphasize the general importance of certain types of training (e.g., education) and certain characteristics (e.g., desired leadership traits) for all leaders in the Army.

The document itself has modernized so that it now includes web links. It also reflects updated timelines for assignments and opportunities as they have evolved within the Army. In sum, the document remains functionally consistent over time but reflects the Army's evolving needs and related requirements.

A6. Term Frequency-Inverse Document Frequency

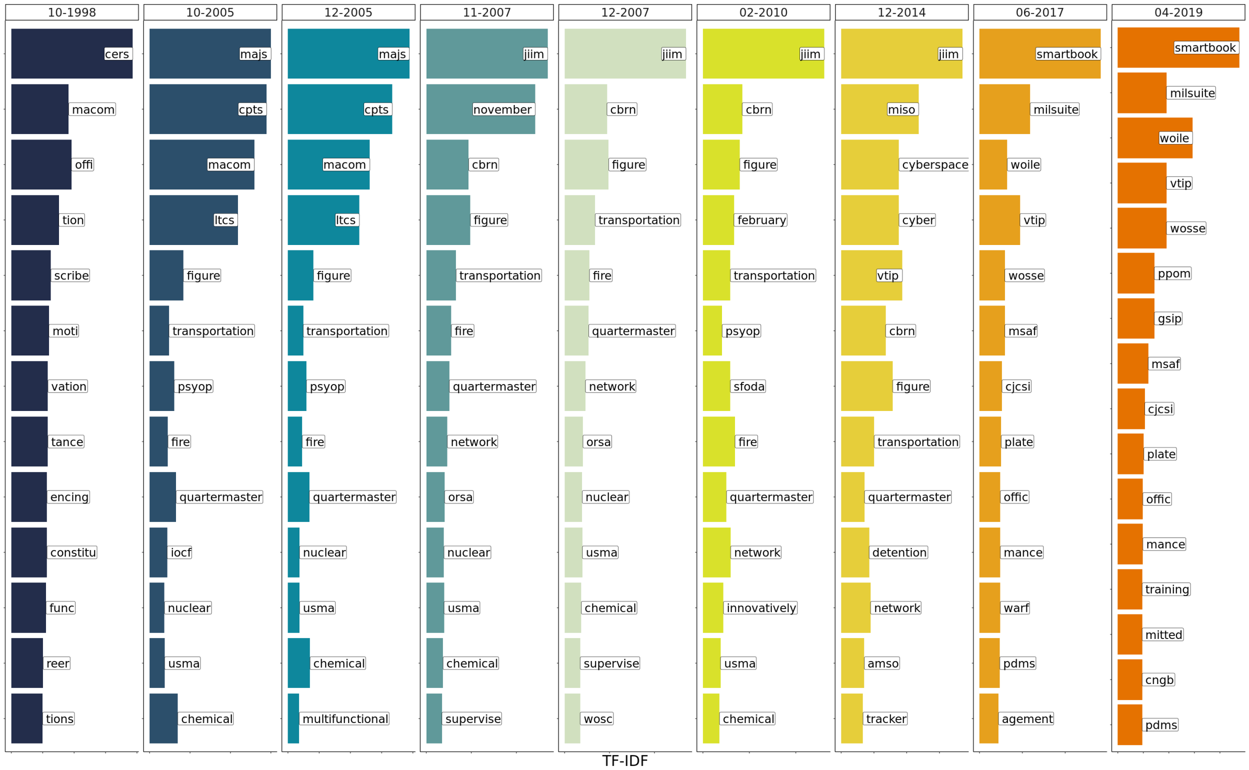
TF-IDF was also used as an analytic approach used to understand apart the changes in the PAM 600-3 over time. The corpus for this analysis (Corpus 2) was nine versions of the PAM 600-3 spanning from 1998-2019:

- U.S. Department of the Army (1998, 2005a, 2005b, 2007b, 2007c, 2010, 2014b, 2017b, 2019d)

Exhibit 52 shows the results of this analysis. We see that some of the most frequent terms that are unique to the documents are acronyms, like JIIM (joint, interagency, intergovernmental, and multinational) and CERS (Consolidated Emergency Response System). We also see program names, like Smartbook. Certain ranks also change in focus over the years, including MAJs (majors) and CPTs (captains). Other topics that change include chemical, nuclear, transportation, and cyber.

Exhibit 53.

Term Frequency-Inverse Document Frequency Results for the PAM 600-3 20 Year Corpus



Note: For the purposes of this study, it is important to note that the colors used for graphs were selected arbitrarily and have no correlation to any derived meaning.

Overall, these results show that there are distinct terms used in the PAM 600-3 over time. As shown in the manual review, however, these distinct terms are not necessarily introducing conceptually distinct ideas in officer career progression. From term frequencies alone, we cannot understand if performance context or concepts in performance are changing.

A7. Sentiment Analysis

To begin to understand conceptual changes in the Corpus 2 (the nine PAM 600-3 documents spanning from 1998-2019), we employed a sentiment analysis. Exhibit 53 shows the results of the sentiment analysis. We find that consistently over time more of the document has a positive sentiment than a negative sentiment. Since 2015, the positive sentiment in the document has grown.

Exhibit 54.
Positive and Negative Sentiment Trends in the Officer Professional Development and Career Management (PAM 600-3) from 1998-2019

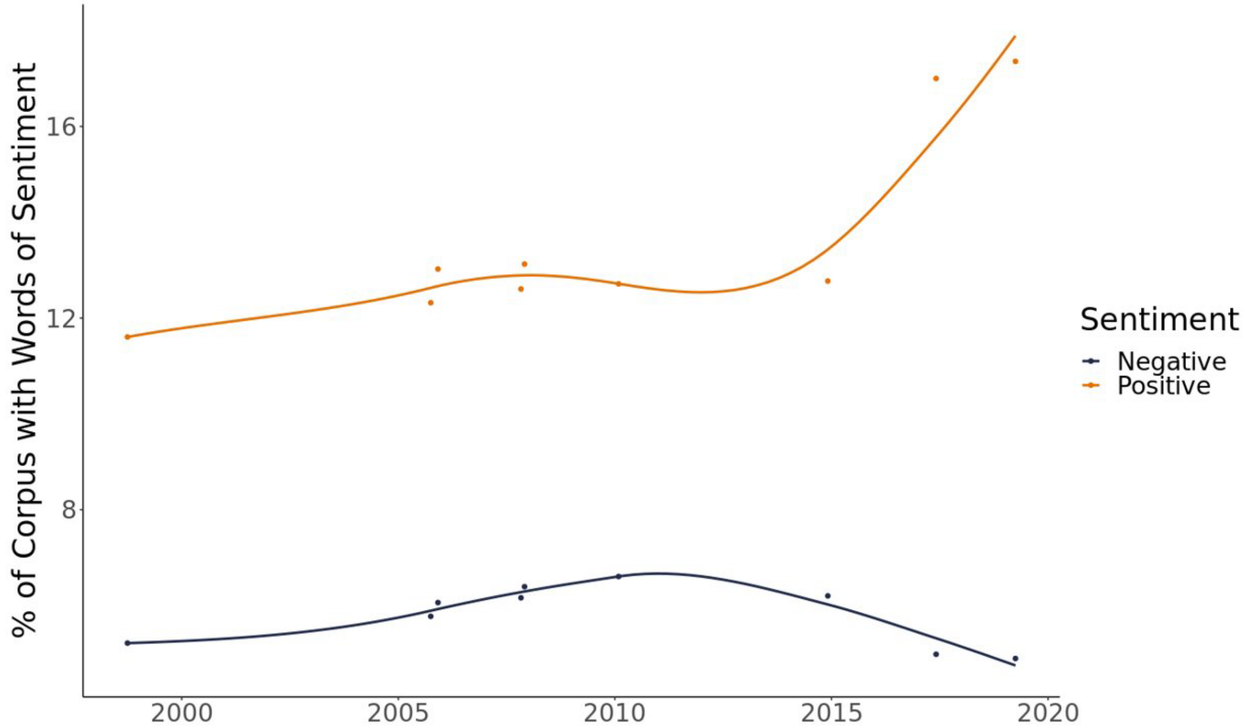
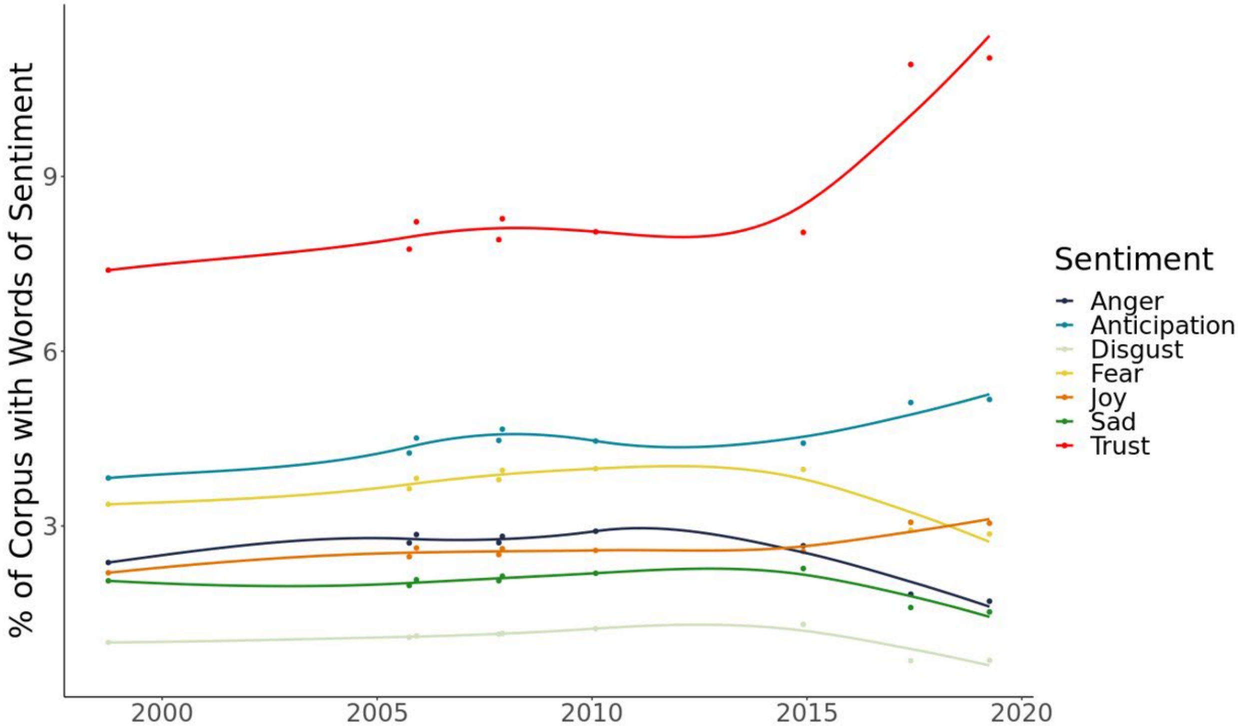


Exhibit 54 **s**hows the results of the detailed sentiment analysis into the emotions driving the change in sentiment of the documents. We see that trust is the dominant emotion of the PAM 600-3 from 1998-2019. Trust has a rapid trend upward since 2015, which is largely driving the positive trend upward. The downward trend seems evenly split among negative sentiment emotions.

Exhibit 55.
Emotional Trends in the Officer Professional Development and Career Management (PAM 600-3) from 1998-2019



These results support our assertion that, conceptually, trust is influential towards Soldier performance. Not only is trust the most prevalent sentiment in the corpus, but it also trends upward in recent years. Given the results of previous analyses as well, the detailed sentiment analysis supports the idea that trust is a Soldier characteristic that contributes to unit performance.

A8. Pronoun Frequency Changes

We also sought to understand changes in the description of performance over time by looking at group and individual pronoun usage, as well as gendered pronoun usage. This analysis also used Corpus 2, the corpus of nine PAM 600-3 documents from 1998 to 2019.

Exhibit 55 shows the changes in individual and collective pronoun usage over time period. We find group-focused pronouns occur more frequently in the PAM 600-3 from 1998-2015. Additionally, the use of both Individually- and collectively-focused pronouns has been increasing since 2015. Interestingly, group-focused pronouns are overtaken by individual pronouns in 2015, and individual pronouns represent proportionally more of the document.

Exhibit 56.
Individual vs Collective Pronoun Trends in the Officer Professional Development and Career Management (PAM 600-3) from 1998-2019

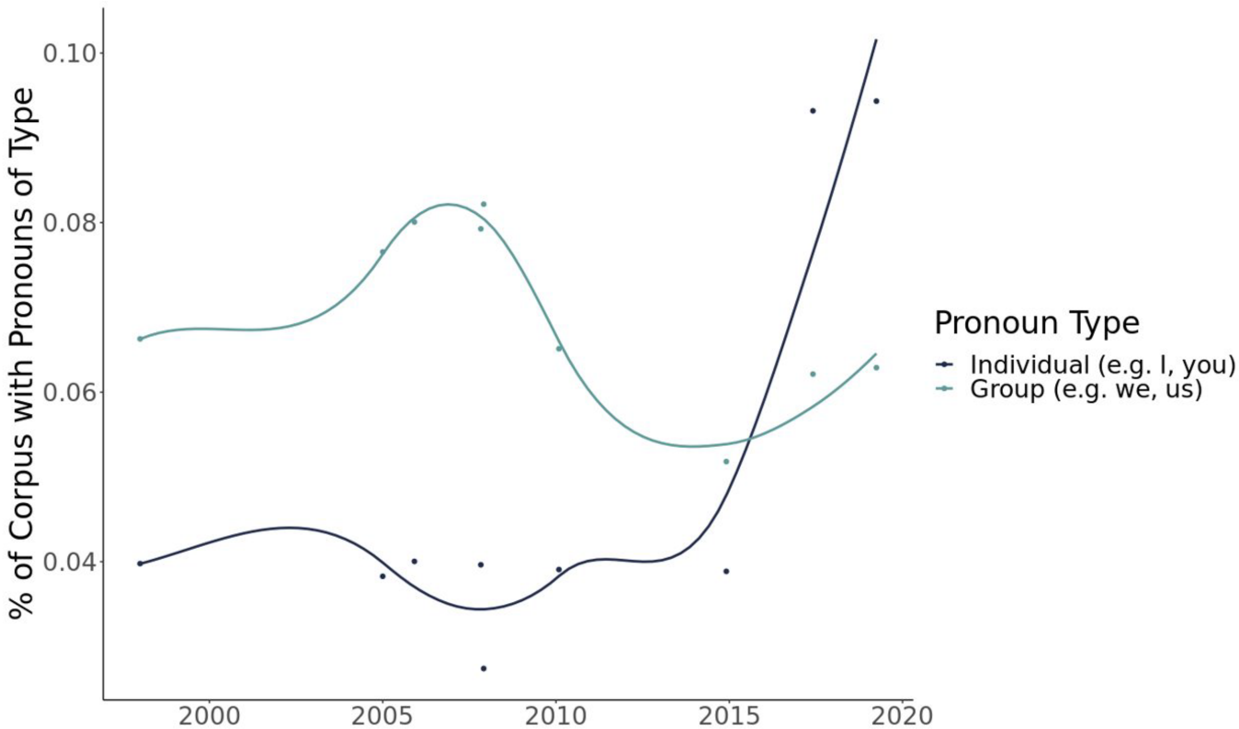
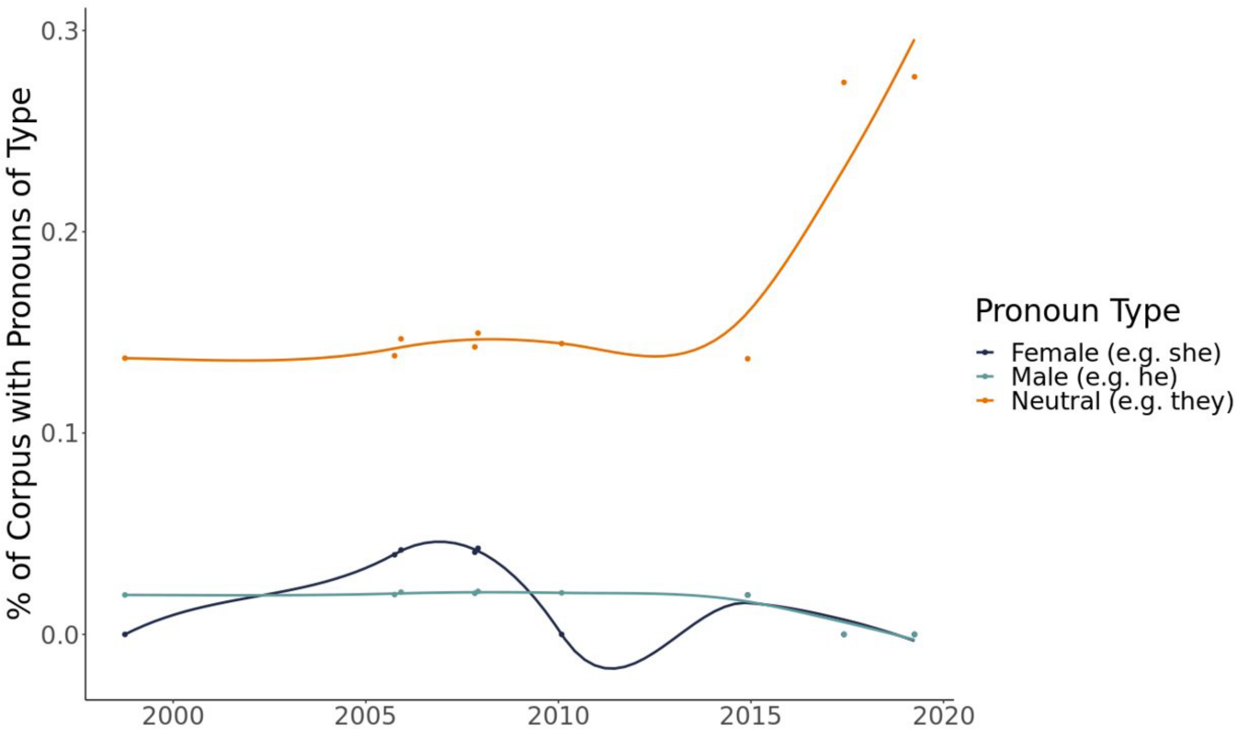


Exhibit 56 shows the changes in gendered pronouns over the 20-year time period. We see that collective-focused (neutral) pronouns occur most frequently in the PAM 600-3 from 1998-2019. The use of neutral, collective pronouns is increasing rapidly since 2015. The use of individually-focused (both male and female) pronouns is decreasing since 2015. The documents between 2005-2007 contain proportionally more female pronouns than male pronouns. In 2010, male pronouns occur more frequently than female pronouns. Since 2010, the proportion of female and male pronouns has stayed even.

Exhibit 57.
Individual vs Collective Pronoun Trends in the Officer Professional Development and Career Management (PAM 600-3) from 1998-2019



Findings from the pronoun frequency analysis suggest that group-focused pronouns are more frequently used in the PAM 600-3. This suggests that groups are a commonly discussed subject in performance publications. Over the time period, we find that since 2015 individual-level pronouns have been relatively more commonly used. The majority of the increase in usage of individual-level pronouns has been driven by neutral (they/them) pronouns, which also can refer to groups.

A9. Using BERT to Discover Changes in Performance Documents Over Time

For A9, we aimed to discover changes in performance documents over time. For this analysis, we used a corpus of Army doctrine that came from the Center for Army Leadership and the Profession (CAPL) (Corpus 3). Importantly, this corpus spanned several decades: from the 1940s until the 2020s. Corpus 2 included:

- U.S. Department of the Army (1946, 1948b, 1948a, 1951, 1953, 1958, 1961, 1965, 1973, 1983, 1985, 1986, 1987c, 1987b, 1987a, 1988, 1990b, 1990a, 1993, 1999a, 1999b, 2002b, 2002a, 2003a, 2003b, 2003c, 2004, 2006, 2007a, 2008, 2009, 2011, 2012a, 2012b, 2014c, 2015c, 2015a, 2016, 2017a, 2019b, 2022)

We aimed to use quantitative analysis to discover trends in these documents over time. Using BERT, we can discover relative changes in these documents. We look at the cosine similarity between documents as well as cosine similarity to baseline documents to discover how the concept of performance or leadership has changed over time.

Exhibit 57 shows the cosine similarity between all documents in the corpus. We find that the most dissimilar documents on average are the FM 22-10 from 1957, the FM 7-22-7 from 2002, and the FM 7-15-C3 from 2010. These documents are all functional manuals, suggesting that lower-level publications (i.e. what is functionally used by Soldiers on the ground) may not reflect entirely what is exposed in Army doctrine. We also find a relatively high cosine similarity between documents since 2012. This indicates that there is relatively more thematic cohesion in the idea of performance and leadership in recent years. We find conceptual cohesion in the overall time frame, given minimum cosine similarity of 0.6.

Exhibit 58.

Cosine Similarity between All Army Performance Related Documents from 1946 to 2021 Using BERT, Ordered by Year

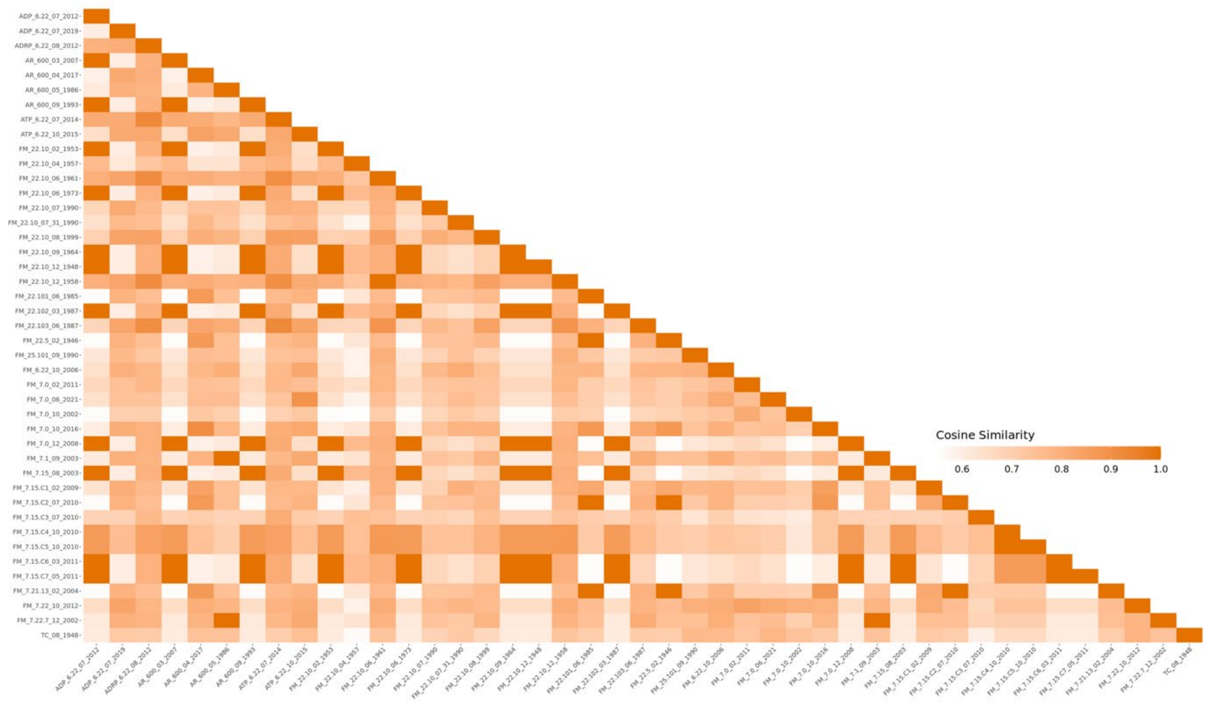
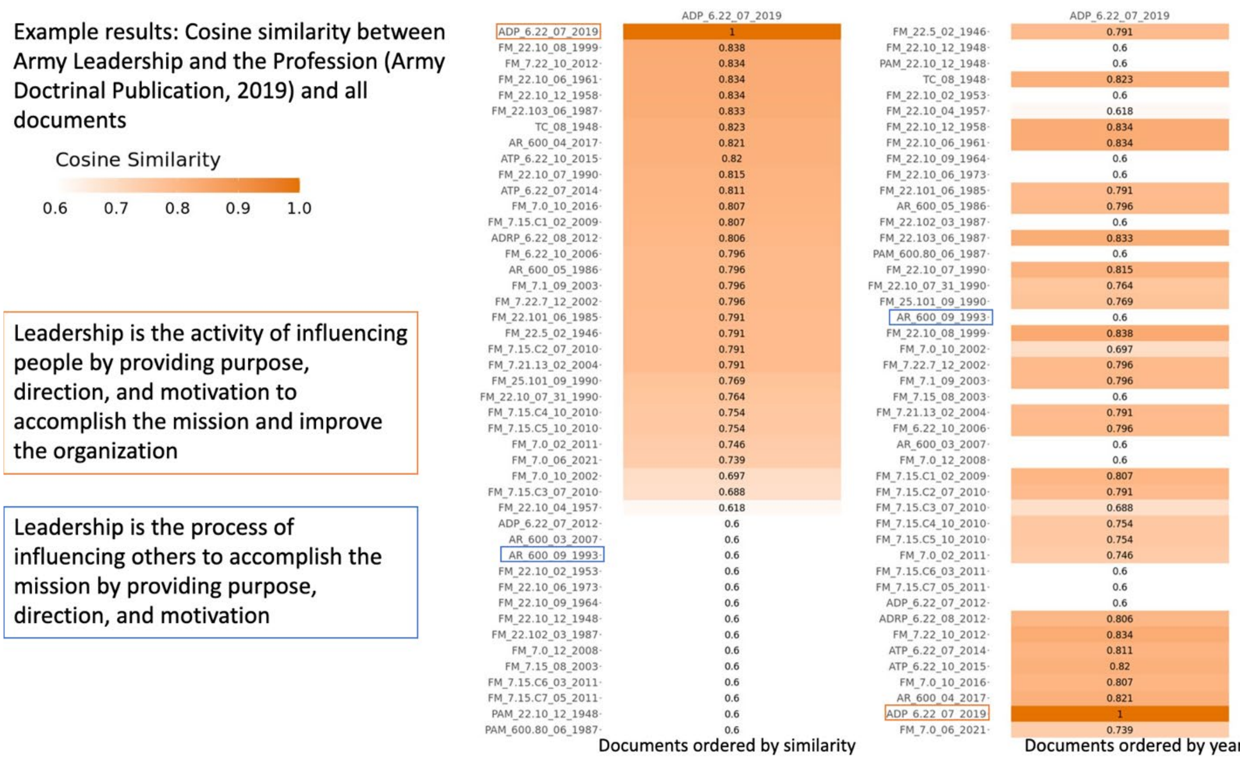


Exhibit 59 shows the cosine similarity between all documents in the corpus and the Army Leadership Development publication (ADP 6-22). We find that, similar to the overall trend, the corpus is relatively more similar to the ADP 6-22 since 2012. As an example, we pull out the definition of leadership according to CAPL in the baseline document and the AR 600-09 from 1993. We find that, although these definitions are semantically similar, these documents have a relatively low cosine similarity. These differences can be attributed to conceptual differences in the documents, which includes 'improving the organization' as a part of the definition, something not included in the AR 600-09 from 1993. It is also important to note (as previously mentioned) that the minimum cosine similarity is 0.6, suggesting that documents do share a degree of conceptual alignment.

Exhibit 59.
Cosine Similarity between ADP 6-22 on Army Leadership Development and All Army Performance Related Documents from 1946 to 2021 Using BERT, Ordered by Similarity (Left) and Year (Right)



RQ4: How well do performance concepts in items assessing Soldier psychosocial social characteristics align with performance concepts in Army publications?

Our next line of inquiry investigated the relationship between survey items and performance. As our project aims to repurpose administrative data, we are interested in discovering performance proxies. We investigated the item text for survey items in the GAT 2.0 to analyze their conceptual alignment with performance documents. Finding alignment may suggest that these variables would be appropriate as inputs in performance modeling efforts. If we do not find alignment, it may suggest that item text could be updated to better align with performance concepts.

A10. Using BERT to Analyze the Convergence of Performance Documents and GAT Items

In A10 we leveraged BERT to discover the conceptual similarity between item text for GAT 2.0 constellation measures and Army Leadership and the Profession (ADP 6-22) (Corpus 4). We found that BERT was not very discerning between performance concepts overall, and less so for productive and counterproductive performance. Exhibit 59 shows the cosine similarity scores for the productive and counterproductive versions of GAT items vectorized using BERT. For example, the cosine similarity score between the positive or productive ("I am extremely satisfied with my marriage") and counterproductive ("I am not at all satisfied with my marriage") item text for the family satisfaction GAT item is 0.92. We find that BERT considers the productive and counterproductive GAT item text to be highly conceptually similar.

Exhibit 60.
Cosine Similarity between the Productive and Counterproductive Characteristic for Each Measure of the GAT

GAT Item	Cosine Similarity
adaptability	1.00
bad coping	1.00
catastrophizing	0.94
character	0.94
depression	0.99
family closeness	0.92
family satisfaction	0.92
family support	0.88
friendship	0.91
good coping	0.91
life meaning	0.92
loneliness	0.9
negative affect	0.92
non-work interests	0.89
optimism	0.89
organizational trust	0.84
positive affect	0.97
work engagement	1.00

Both productive and counterproductive psychosocial characteristics were conceptually represented in the performance document according to BERT. Exhibit 60 shows the cosine similarity between each measure of the GAT and our benchmark performance document, the ADP 6-22. For example, the cosine similarity score between the positive or productive ("I am extremely satisfied with my marriage") item text for the family satisfaction GAT item and the ADP-622 is 0.91.

- The following characteristics aligned most with the performance document.
    - Negative affect - counterproductive performance (e.g. "I often feel upset")
    - Friendship - counterproductive performance (e.g. "I do not have a best friend")
    - Character - productive performance (e.g. "In the last four weeks I have often acted with honesty.")
    - Non-work interests - productive performance (e.g. "It is very much like that my work is one of the most important things in my life")
    - Work engagement - productive performance (e.g. "It is very much like me at all that my work is one of the most important things in my life")
- The following characteristics aligned least with the performance document.
    - Optimism - productive performance (e.g. "It is very much like me that in uncertain times, I usually expect the best")
    - Family satisfaction - counterproductive performance (e.g. "In the past four weeks I have felt not at all satisfied with my family")

Exhibit 61.
Cosine Similarity between Each Measure of the GAT and the ADP 6-22

Productive GAT Item	Cosine Sim.	Counterproductive GAT Item	Cosine Sim.
adaptability	0.99	adaptability	0.99
bad coping	0.91	bad coping	0.92
catastrophizing	0.90	catastrophizing	0.90
character	1.00	character	0.99
depression	0.90	depression	0.94
family closeness	0.99	family closeness	0.91
family satisfaction	0.91	family satisfaction	0.86
family support	0.89	family support	0.91
friendship	0.91	friendship	1.00
good coping	0.89	good coping	0.91
life meaning	0.89	life meaning	0.89
loneliness	0.92	loneliness	0.99
negative affect	0.89	negative affect	1.00
non-work interests	1.00	non-work interests	0.89
optimism	0.84	optimism	0.99
organizational trust	0.89	organizational trust	0.89
positive affect	0.91	positive affect	0.91
work engagement	1.00	work engagement	0.91

A11. Using GPT-2 to Analyze the Convergence of Performance Documents and GAT Items

We also analyzed Corpus 4 using GPT-2. We found that GPT-2 was even less discerning between performance concepts than BERT. In other words, cosine similarity scores between GAT items and the ADP-622 were consistently very high, indicating indiscriminately high conceptual similarity.

RQ5: What descriptive and prescriptive behaviors are associated with high performing Soldiers?

A12. Term Frequency in Army Task Lists

In A12 we performed a term frequency, extracting verbs specifically, on the Soldier's Manual of Common Tasks: Warrior Skills Level 1 (STP 21-1-SMCT) (U.S. Department of the Army, 2021) (Corpus 6). Overall, 13.2% of words in the Soldier's Manual of Common Tasks are tagged as verb part of speech. This accounted for 1,179 unique verbs. The most frequently mentioned verbs were:

- Use (435 occurrence)
- Figure (270 occurrences)
- Go (245 occurrences)
- Identify (214 occurrences)
- Do (202 occurrences)

A13. Verb Phrases (Co-occurrence) in Army Task Lists

For A13 we analyzed term co-occurrence using Corpus 6. We found that frequently co-occurring words contextualized behaviors in the Soldier's Manual of Common Tasks. For example, "use" frequently cooccurs with:

- People e.g. ("Soldier", "student")
- Objects e.g. ("map", "compass")
- Qualifiers e.g. ("without", "must")

Overall, we observed that frequently occurring behaviors seem to connect to leadership activities generally, such as communication, rather than warfighting. We found that these behaviors frequently co-occurred with the term rank associated terms, indicating that hierarchies or communication vertically is an important performance behavior.

Summary

Overall, we find several topics relevant to Soldier performance. Identification of relevant topics, particularly those that can be represented by data in the PDE. The document analysis provides a backdrop for results of performance models. We find that the way the Army has conceptualized performance has shifted over time but has ultimately remained very similar. This is supported by internal Army sources, including CAPL. We also find that psychosocial characteristics and performance behavior are somewhat difficult to measure using LLMs. In particular, LLMs are not discerning between productive performance and counterproductive performance, showing relatively high conceptual cohesion overall. This could be the result of the item text itself but, more likely, it is due to the generalized training datasets of large language models and their aptitude for detection of global over local contextual information.

PART 4: SUMMARY AND CONCLUSION

In this final report, we summarize our research on the history of measuring performance in the Army, present an overview of the performance literature, and create a conceptual model that sets the stage for our quantitative and qualitative analyses. This is followed by an extensive analysis of the PDE data through conceptual and methodological profiling, descriptive analysis of the data over 20 years, and a simple analysis of the data. Seven variables are identified representing a spectrum of negative and positive outcomes (alcohol use, bad paper counts, first term attrition, physical fitness scores, award counts, speed of promotion, and character of service). An example of a simple logistic regression model to test these outcomes is provided. We conclude with a mixed model analysis involving qualitative and natural language processing techniques that finds that the way the Army has described performance has shifted over time but has ultimately remained very similar.

This research highlights the possibilities of measuring performance using administrative and survey data already collected by the Army – for example, measuring the psychosocial traits from the Global Assessment Tool (GAT), now called the Azimuth survey – and connecting them to outcomes. Still, there are limitations in working with administrative data to address complex, nuanced concepts like performance. For example, task performance is not captured very well in administrative data. However, the data do paint a picture of key aspects of the social dimension, for example a Soldier's willingness to fight, which can be identified and measured in data. Importantly, we show that these factors speak to key performance outcomes, like speed to promotion, even without including task-specific data.

The qualitative analysis was important in understanding how the Army thinks about performance over time. Soldier's jobs will change with time as technology changes the specific skills, they need to do their jobs, but the Army's mission remains the same. The qualitative analysis reinforced that understanding and measuring a Soldier's attitude and willingness to engage in their work are also constant elements.

This project has added value to the performance literature by systematically capturing and measuring the quantitative and qualitative information the Army produces and linking it to the social components of performance outcomes.

References

Adjerid, I., & Kelley, K. (2018). Big data in psychology: A framework for research advancement. *American Psychologist*, *73*, 899-917. https://doi.org/10.1037/amp0000190

ADP 1-01: Doctrine Primer. (2019). Department of the Army.

Azen, R., & Budescu, D. V. (2003). *The dominance analysis approach for comparing predictors in multiple regression*. Psychological Methods, 8, 129-148. https://doi.org/10.1037/1082-989x.8.2.129

Azen, R., & Budescu, D. V. (2006). Comparing predictors in multivariate regression models: An extension of dominance analysis. *Journal of Educational and Behavioral Statistics*, *31*, 157–180. https://doi.org/10.3102/10769986031002157

Berman, F., Rutenbar, R., Hailpern, B., Christensen, H., Davidson, S., Estrin, D., Franklin, M., Martonosi, M., Raghavan, P., Stodden, V., & Szalay, A. (2018). Realizing the potential of data science. *Communications of the ACM*, 61(4), 67-72. https://dl.acm.org/citation.cfm?doid=3200906.3188721

Borman, W.C., Johnson, P.D., Motowidlo, S.J., & Dunnette, M.D. (1975). Measuring motivation, morale, and job satisfaction in Army Careers. Technical Report #716 (ARI Research Note 79-30). U.S. Army Research Institute for the Behavioral and Social Sciences, USA.

Borman, W.C., Motowidlo, S.J., Rose, S.R., & Hanser, L.M. (1987). Development of a Model of Soldier Effectiveness. Technical Report #741 (ARI Research Note 87-29). U.S. Army Research Institute for the Behavioral and Social Sciences, Arlington, VA.

Budescu, D. V. (1993). Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin*, 114, 542-551. https://doi.org/10.1037/0033-2909.114.3.542

Campbell, J. P., & Knapp, D. J. (Eds.). (2001). *Exploring the limits in personnel selection and classification.* Lawrence Erlbaum Associates Publishers.

Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1993). A Theory of Performance. In Schmitt, N. Borman, W., and Associates (Eds.). *Personnel Selection in Organizations* (pp. 35-70). Jossey-Bass Publishers, San Francisco.

Campbell, J. P., & Wiernik, B. M. (2015). The modeling and assessment of work performance. *The Annual Review of Organizational Psychology and Organizational Behavior*. 2, 47-74. doi: 10.1146/annurev-orgpsych-032414-111427

Cox, M. (2019, April 24). Army's New Promotion System Could Mean Separations for Some Soldiers. *Military.Com.*

Data Science for the Public Good. (2020, August 7). *Army Performance Measurement: Content and Themes*. https://dspg-young-scholars-program.github.io/DSPG2020_ARI2/?dspg

Department of Defense. (2003). *2003 Demographics: Profile of the military community*. Washington, DC. https://prhome.defense.gov/M-RA/Inside-M-RA/MCFP/Reports/ [September 2020]

Department of Defense. (2010). *2010 demographics report: Profile of the military community*. Washington, DC. https://download.militaryonesource.mil/12038/MOS/Reports/2010-Demographics-Report.pdf [September 2020]

Department of Defense. (2018). *2018 demographics report: Profile of the military community*. Washington, DC. https://download.militaryonesource.mil/12038/MOS/Reports/2018−demographics−report.pdf [September 2020]

Department of Defense. (2019). *2019 demographics report: Profile of the military community*. Washington, DC. https://download.militaryonesource.mil/12038/MOS/Reports/2019-demographics-report.pdf [September 2020]

Doan, A., Halevy, A., & Ives, Z. (2012). Data provenance. *Principles of data integration* (pp. 359–371). Morgan Kaufmann. https://doi.org/10.1016/B978-0-12-416044-6.00014-4

Fischl, M.A., Gorman, S.H., Hoshaw, R.R., Maier, C.R., & Milton H. (1980). History of the Armed Services Vocational Aptitude Battery (ASVAB) 1974-1980. Technical Report. Armed Services Vocational Aptitude Battery Working Group, USA.

Garber, A. M. (2019). Data Science: What the Educated Citizen Needs to Know. *Harvard Data Science Review, 1.1,* 1-14. https://doi.org/10.1162/99608f92.88ba42cb

Gaucher, D., Friesen, J., & Kay, A. C. (2011). Evidence that gendered wording in job advertisements exists and sustains gender inequality. *Journal of Personality and Social Psychology*, *101*(1), 109-128. https://doi.org/10.1037/a0022530

General Records of the United States Government; Record Group 11. (1948). *Executive Order 9981*.

Glavic, B., & Dittrich, K. R. (2007). Data provenance: A Categorization of existing approaches. In: Kemper, A., Schöning, H., Rose, T., Jarke, M., Seidl, T., Quix, C., & Brochhaus, C. (Eds.), *Datenbanksysteme in Business, Technologie und Web (BTW) / 12. Fachtagung des GI-Fachbereichs Datenbanken und Informationssysteme (DBIS)* (pp. 227–241). Aachen. Bonn.

Google. (2023, August 8). *Introduction to Large Language Models*. Google for Developers.

Gunjal, S. (2020). *Tokenization in NLP*. Kaggle.

Headquarters, Department of the Army. (2019). Army Leadership and the Profession (ADP 6-22).

Headquarters, Department of the Army. (2012). Training Units and Developing Leaders (ADRP 7-0).

Howell, D. C. (2008). *Fundamental Statistics for the Behavioral Sciences* (6th ed.). Thomson Wadsworth.

Joint Chiefs of Staff. (2017). Joint Operations (JP 3-0). Washington, DC.

Keller, S., Korkmaz, G., Robbins, C., & Shipp, S. (2018). Opportunities to observe and measure intangible inputs of innovation: Definitions, operationalization, and examples. *PNAS*, *115*, 12638-12645. https://doi.org/10.1073/pnas.1800467115

Keller, S. A., Shipp, S. S., Schroeder, A. D., & Korkmaz, G. (2020). Doing Data Science: A Framework and Case Study. *Harvard Data Science Review*, 2(1). https://doi.org/10.1162/99608f92.2d83f7f5

King, E. B., Tonidandel, S., Cortina, J. M., & Fink, A. A. (2016). Building understanding of the data science revolution and I-O psychology. In: Tonidandel, S., King, E. B., & Cortina, J. M. (Eds.), *Big Data at Work: The Data Science Revolution and Organizational Psychology* (pp. 1-15). Routledge. https://doi.org/10.4324/9781315780504

Kline, R. B. (2011). *Principles and Practice of Structural Equation Modeling* (5th ed., pp. 3–427). The Guilford Press.

Koopmans, L., Bernaards, C. M., Hildebrandt, V. H., Schaufeli, W. B., de Vet Henrica, C. W., & van der Beek, A. J. (2011). Conceptual frameworks of individual work performance: a systematic review. *Journal of Occupational and Environmental Medicine*, 53(8), 856-866.

Knapp, D., Asch, B. J., DeMartini, C., Ruder, T., & Hanley, J. M. (2018a). *Using the Person-Event Data Environment for military personnel research in the department of defense: An evaluation of capability and potential uses*. RAND Corporation: Santa Monica, CA., https://www.rand.org/pubs/research_reports/RR2302.html

Knapp, D., Orvis, B. R., Maerzluft, C. E., & Berglund, T. (2018b). *Resources required to meet the U.S. Army's enlisted recruiting requirements under alternative recruiting goals, conditions, and eligibility policies*. RAND Corporation: Santa Monica, CA.

Lord, F. M. (1953). On the statistical treatment of football numbers. *American Psychologist*, 8, 750–751. https://doi.org/10.1037/h0063675

Miner, A. S., Stewart, S. A., Halley, M. C., Nelson, L. K., & Linos, E. (2023). Formally comparing topic models and human-generated qualitative coding of physician mothers' experiences of workplace discrimination. Big Data & Society, 10(1). https://doi.org/10.1177/20539517221149106

National Academies of Sciences, Engineering, and Medicine. (2017). *Strengthening Data Science Methods for Department of Defense Personnel and Readiness Missions*. National Academies Press.

National Research Council (US) Committee on Responsibilities of Authorship in the Biological Sciences. (2003). The Purpose of Publication and Responsibilities for Sharing. In *Sharing Publication-Related Data and Materials: Responsibilities of Authorship in the Life Sciences*. National Academies Press (US).

O'Connell, D. (n.d.). *The Truth About Promotional Review Boards*.

Pulakos, E.D. & Borman, W.C. (1986). Development and Field Test of Army-Wide Rating Scales and the Rater Orientation and Training Program. Technical Report #716. U.S. Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA.

Pulakos, E.D., Borman, W.C., & Hough, L.M. (1988). Test validation for scientific understanding: Two demonstrations of an approach to studying predictor-criterion linkages. *Personnel Psychology*, 41, 703-716.

Provost, F. & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data*, *1*, 51-59. https://doi.org/10.1089/big.2013.1508

Ratcliff, N. & Thurston, J. (2022) Conceptual and Methodological Profiling of Data: A Qualitative Metadata Process to Understand the Meaning and the Formation of Data Sources, Technical Report. TR# 2023-284. Proceedings of the Biocomplexity Institute, University of Virginia; January. DOI: https://doi.org/10.18130/we5y-ez98.

Ratcliff N, Thurston J, Linehan K, Mikytuck A, Oh E, Halder A, Goldstein J, Lancaster V, Schroeder A, Shipp S, Keller S. (2021) A 20-Year Review of Accession Characteristics in the U.S. Army, (unpublished). Technical Report BI-2023-285. Proceedings of the Biocomplexity Institute, University of Virginia; August.

Ratcliff, N., Thurston, J., Goldstein, J., Lancaster, V., Shipp, S., Keller, S., & Ervin, K. (2021a). Examining the population level and individual level longitudinal stability of psychosocial measures in the U.S. Army's Global Assessment Tool (GAT). *Military Psychology*. https://doi.org/10.1080/08995605.2021.1984741

Ratcliff, N., Thurston, J., Goldstein, J., Lancaster, V., Shipp, S., Keller, S., & Ervin, K. (2021b). An Examination of the Longitudinal Stability of Psychological Measures Contained within the U.S. Army's Global Assessment Tool. Technical Report (#1397). U.S. Army Research Institute for the Behavioral and Social Sciences, USA. https://apps.dtic.mil/sti/trecms/pdf/AD1147839.pdf

Ratcliff, N., Thurston, J., Goldstein, J., Halder, A., Mikytuck, A., Lancaster, V., Schroeder, A., Shipp, S., & Keller, S. (2021c). *Leveraging existing administrative data to predict individual performance in the Army (Modeling Phase 1 Technical Report): Simple relationship analysis*. (unpublished). Technical Report BI-2021-139. Proceedings of the Biocomplexity Institute, University of Virginia; August.

Ratcliff, N., Thurston, Linehan, K., J., Goldstein, J., Lancaster, V., Schroeder, A., Shipp, S., & Keller, S. (unpublished). Benchmarking Crosswalks of Army data from Multiple Sources.

Rumsey, M. G. (2012) Military Selection and Classification in the United States. In J. H. Laurence & M. D. Matthews (Eds.), *An International Perspective on Military Psychology*. Oxford University Press, USA.

Sackett, P. M., Lievens, F., Van Iddekinge, C. H., & Kuncel, N. R. (2017). Individual differences and their measurement: A review of 100 years of research. *Journal of Applied Psychology*, 102(3), 254-732.

Schmitt, E., & Helene, C. (2021, April 6). Promotions for Female Generals Were Delayed Over Fears of Trump's Reactions. *The New York Times*.

Simmhan, Y. L., Plale, B., & Gannon, D. (2005). A survey of data provenance in e-science. *ACM SIGMOD Record*, *34*, 31–36. https://doi.org/10.1145/1084805.1084812

Song, J., Alter, G., & Jagadish, H. V. (2019). C2Metadata: Automating the capture of data transformations from statistical scripts in data documentation. *In Proceedings of the 2019 International Conference on Management of Data (SIGMOD '19). Association for Computing Machinery*. https://doi.org/10.1145/3299869.3320241

Steinhauer, J. (2021, April 6). The Army is Adjusting its new fitness test in response to criticism. *The New York Times*.

Tang, S. (2014). *Inferential procedures for dominance analysis measures in multiple regression*. Theses and Dissertations. 645.

Thulin, L. (2019). The First Personality Test Was Developed During World War I. Smithsonian Magazine. https://www.smithsonianmag.com/history/first-personality-test-was-developed-during-world-war-i-180973192/

U.S. Army Training and Doctrine Command (TRADOC). (2011). The U.S. Army Learning Concept for 2015 (PAM 525-8-2).

van der Aalst, W. M. P. (2020). The data science revolution: How learning machines changed the way we work and do business. In: Strous, L., Johnson, R., Grier, D. A., & Swade, D. (Eds.), *Unimagined futures: ICT opportunities and challenges* (pp. 5-19). Springer. https://doi.org/10.1007/978-3-030-64246-4_2

Vie, L. L., Scheier, L. M., Lester, P. B., Ho, T. E., Labarthe, D. R., Seligman, M. E. P. (2015). The U.S. Army Person-Event Data Environment: A military-civilian big data enterprise. Big Data, 3, 1-13. https://doi.org/10.1089/big.2014.0055

Vindman, A. (2021, February 18). *Situational Ethics Drove the Military's Decision to Delay Female Officer Promotions*. The New York Times.

White, L.A., & Young, M.C. (1998). Development and validation of the Assessment of Individual Motivation (AIM). Paper presented at the annual meeting of the American Psychological Association, San Francisco, CA.

Wing, J. M., Janeja, V. P., Kloefkorn, T., & Erickson, L. C. (2018). Data Science Leadership Summit: Summary Report. Technical Report. National Science Foundation, USA. https://doi.org/10.13140/RG.2.2.13710.61764

Yan, X., Guo, J., Lan, Y., & Cheng, X. (2013). *A Biterm Topic Model for Short Texts*. International World Wide Web Conference Committee (IW3C2), Rio Ode Karo, Brazil. https://github.com/xiaohuiyan/xiaohuiyan.github.io/blob/master/paper/BTM-WWW13.pdf