Breaking Silos to Build the Next Generation of Navy Leaders

LCDR Michael W. Natali, Ph.D.
Office of Naval Research
Arlington, VA
michael.w.natali.mil@us.navy.mil

John S.W. Spinda, Ph.D., Bradley D. Celestin, Ph.D., Holly C. Baxter, Ph.D.
Cognitive Performance Group
Portsmouth, VA
john@cognitiveperformancegroup.com,
brad@cognitiveperformancegroup.com,
holly@cognitiveperformancegroup.com

ABSTRACT

The U.S. Navy is developing a Navy Leadership Behavioral Model (NLBM) to improve assessment of leadership potential, develop effective leadership across Sailors, and select the right leaders for the right positions. Organizations traditionally rely on theory-driven leadership models with limited empirical validation. However, the Navy is collecting extensive data to empirically derive and validate a NLBM across Navy communities. This will set the foundation for a comprehensive talent management system to track and support personalized Sailor assessment, career-long leadership development, and command selection via an enterprise-wide software application connecting disparate data sources. This paper will describe the development of the NLBM and supporting empirical evidence.

A multi-stage, thematic analysis identified approximately 20 behavioral factors illustrating effective Navy leadership. First, a qualitative sorting of raw interview data from five Navy commands (n = 1,689) was used to group related statements; these were refined into an initial set of behavioral factors. Second, additional interview data was collected for validation and subsequent thematic coding (n = 2,131). Concurrently, two workshops with Navy subject matter experts were held to refine the behavioral factors and their operational definitions. Leveraging the refined model, all qualitative data was combined into a single dataset for a comprehensive sorting and quantitative analysis. This last step allowed us to assess the (1) metrics of inter-coder reliability, (2) factor commonalities via natural language processing analysis, and (3) Navy community priorities to support the current NLBM structure.

Because organizations rarely validate leadership behavioral models, sharing the empirical process for developing the NLBM will demonstrate how the combination of well-known and novel methods for analyzing qualitative data can provide stronger and more relevant solutions than relying on theory alone. Additionally, the paper will discuss how conducting the empirical analysis supports long-term validation against relevant criteria and advanced data analytics across disparate data sources.

ABOUT THE AUTHORS

LCDR Michael W. Natali, Ph.D., is the Manpower, Personnel, Education, & Training Program Officer for Office of Naval Research where he oversees innovative research expanding Navy's capability to select, man, and train Sailors. Additionally, he has managed research, development, testing, and evaluation for programs enhancing aviator performance, identifying emerging technologies to integrate into aircraft systems, training, and processes ensuring the continued dominance of Naval aviation. LCDR Natali completed his doctorate in I/O Psychology at University of Minnesota while also working as a Research Consultant for Assessment Associates International. He joined the Navy in 2014 as an Aerospace Experimental Psychologist and his expertise covers personnel selection, training systems and evaluation, performance measurement, and human-systems integration within military settings.

Holly C. Baxter, Ph.D., Chief Scientist of Cognitive Performance Group, has spent over 25 years specializing in cognitively based Instructional Design, Evaluation Metrics, Organizational Development, and Training in both military and commercial environments. Dr. Baxter has published numerous articles in the fields of cognitively-based training, assessment, and knowledge management; has been an invited speaker at multiple conferences and events; and has given dozens of workshops on Cognitive Task Analysis, Knowledge Capture & Transfer, Knowledge Management Assessment, Intuitive Decision-

Making, and Leadership Development. Dr. Baxter earned a Ph.D. from Indiana University in Organizational Communication and Human Resource Management.

John S. W. Spinda, Ph.D., is a Scientist with Cognitive Performance Group where he uses a portfolio of broad-based research skills to lead and support government-funded, operational research designed to improve leader effectiveness, communication skills, and data visualization. His capabilities include quantitative and qualitative research methods and analysis, competency modeling, scale development, learning assessments, as well as instructional design and delivery. In addition to his industry work, John spent 15 years in higher education, instructing a broad and diverse catalog of courses and conducting communication-based research in media use and social psychology, earning top paper awards and peer-reviewed publications.

Bradley D. Celestin, Ph.D., is a Senior Scientist at Cognitive Performance Group, with expertise in social, cognitive, quantitative, and computational psychology. He previously served as the chair of psychology and philosophy and director of software development at Bethel College in North Newton, KS. Dr. Celestin's experience includes curriculum development, teaching, public speaking, coding, research management, data analysis, software design, and investigative interviewing. He has done research in legal decision-making, moral cognition, teaching assessment, and statistical methods. Prior to academia, he spent over a decade as a law enforcement officer specializing in technology-facilitated crime at both municipal and federal levels. Dr. Celestin holds a Ph.D. in Psychology from Indiana University, Bloomington.

Breaking Silos to Build the Next Generation of Navy Leaders

LCDR Michael W. Natali, Ph.D.
Office of Naval Research
Arlington, VA
michael.w.natali.mil@us.navy.mil

John S.W. Spinda, Ph.D., Bradley D. Celestin, Ph.D., Holly C. Baxter, Ph.D.
Cognitive Performance Group
Portsmouth, VA
john@cognitiveperformancegroup.com,
brad@cognitiveperformancegroup.com,
holly@cognitiveperformancegroup.com

INTRODUCTION

The United States (U.S.) Navy is developing a Navy Leadership Behavioral Model (NLBM) as part of an expansive effort to modernize talent management. The primary aims of the NLBM are to improve assessment of leadership potential, train effective leadership behaviors throughout Sailors' careers, and help ensure the right leaders are retained and slated for the leadership positions that best use their abilities. The NLBM is designed to reflect leadership behaviors Navy-wide at the major command and department head levels (i.e., E-7 and up, O-5 and up), such as command triad teams, consisting of a Commanding Officer, Executive Officer, and Senior Enlisted Leader. Also, the NLBM is intended to serve as the foundation of a comprehensive system being built to track individual Sailors' career-long leadership development using an enterprise-level, profile-based technology solution that links disparate databases of information across the Navy. This solution will provide Sailors a clear sense of what they have achieved as leaders and where they can improve to become better leaders.

Presently, the U.S. Navy does not have a force-wide leadership model that distinguishes the behaviors of exemplary leaders. Instead, many individual Navy communities (e.g., Naval Surface Forces, Submarine Forces, Explosive Ordinance Disposal) have developed their own models of leadership based on internal reviews regarding the behaviors reflective of effective leadership within their community. This siloed approach is not without merit; Navy communities fulfill a wide and diverse range of roles and missions to achieve objectives. However, other branches of the U.S. military with numerous communities and varying roles, such as the Army (Department of the Army, 2022) and Air Force (Department of the Air Force, 2022), have successfully adopted force-wide leadership models. In addition, a siloed approach to leadership modeling runs contrary to the evolving and collaborative nature of modern naval leadership. This has been highlighted in recent Navy reports and initiatives (e.g., Department of the Navy, 2019; 2021; 2022). Therefore, the NLBM aims to define effective leadership Navy-wide while still working in concert with existing leadership models that have been developed within Navy communities.

Leadership Models: One Size Does Not Fit All

Numerous academic and organizational researchers have developed an extensive collection of comprehensive leadership models that outline effective (or ineffective) leader behaviors that occur across different types of organizations. For instance, Yukl et al. (2002) created a model of 12 specific leadership behaviors culled from a review of decades of research on leadership. These behaviors were grouped into the meta-factors of task behavior, change behavior, and relations behavior. Avolio and Bass's (1991) Full Range Leadership Model parses 12 leadership characteristics into a set of three leadership types (transformational, transactional, passive-avoidant) that describe high, moderate, and low levels of effort and engagement from leaders. Irrespective of the terms used across this deep pool of comprehensive leadership models, most have a few key commonalities (Campbell, 2012a). These commonalities include the use of three or four meta-factor solutions with each meta-factor containing factors comprised of multiple specific behaviors or competencies.

Comprehensive models have utility because they condense volumes of leadership research into one product. But no matter how accurately a leadership model depicts leadership *across* organization types, it does not assure that the model is a good option for a *specific* organization. Empirical validation along with an actionable plan to use the model to improve leadership (e.g., assessment, development, selection) is a necessary step that organizations often bypass (McLean & Moss, 2003). One example of a tailored approach to model development is 3M's leadership competency

model (Alldredge & Nilan, 2000). Rather than imposing a singular comprehensive leadership model, 3M opted to use existing leadership models as a starting point for developing a model that fit the culture and history of 3M. The result was a model that aligns 12 leadership competencies for company executives into fundamental, essential, and visionary categories. This model was subsequently used for talent management efforts at 3M, such as leader assessment and selection.

Similarly, an effective model of leadership behaviors for the U.S. Navy should account for the nature, history, and culture of the organization while also incorporating the volumes of knowledge generated by more than a century of leadership research: a synthesis of the leadership research literature with Navy-specific data. Such a model should describe leadership behaviors Navy-wide but also have customizable options for individual Navy communities.

NLBM DEVELOPMENT

The NLBM was developed using findings from raw interview and focus group data to develop a model to fit the needs of the U.S. Navy while being anchored in the research literature. More specifically, transcripts from semi-structured interviews and focus groups were provided that represented insights from six different Navy communities (n = 3,820; Cigularov & Navy Expeditionary Combat Command, 2021; Naval Submarine Medical Research Laboratory, n.d.-a, n.d.-b, n.d.-c; Naval Surface Warfare Center Dahlgren, n.d.-a; n.d.-b, 2024). These communities included Naval Air Forces (AIRFOR), Engineering Duty Officer (EDO), Explosive Ordinance Disposal (EOD), Human Resources (HR), Submarine Forces (SUBFOR), and Naval Surface Forces (SURFOR). These data were collected by multiple research teams using different sets of interview questions and prompts. However, each interview or focus group predominantly involved participants being asked to describe some combination of characteristics, practices, and specific instances of both effective and ineffective leadership that they have experienced while in the Navy. The current version, NLBM 3.0, was iteratively developed with periodic new data, SME coding and reviews, and statistical analysis.

Qualitative Card Sort

In the first phase of the analysis after several interview and focus group data collection events, three researchers simultaneously conducted a qualitative card sort to group related responses into categories that reflected similar themes or characteristics of naval leadership. This process was completed over three days using the available raw data at the time of analysis (n = 1,689). Researchers examined each response from the interviews and focus groups on separate slips of paper, stopping at intervals to describe emerging categories by writing a short description of the theme (e.g., “stress management”) or characteristic (e.g., “builds trust”) on an index card. The card sort process originally resulted in 64 categories of leadership characteristics. To distill these into a more manageable framework, further evaluation of the categories merged categories into common themes supported by the literature or removed them because of redundancy or infrequent occurrence. This process resulted in 20 leadership factors grouped across five meta-factors: the initial NLBM 1.0 draft.

Subject matter experts (SMEs) and project stakeholders from a variety of Navy communities and other branches of the U.S. military participated in a two-day workshop to provide feedback on the NLBM 1.0 draft. Attendees proposed revisions, deletions, and changes to meta-factor titles, factor titles, and operational definitions. The final version of the NLBM 1.0 contained 18 leadership factors across five meta-factors: Character, Command Climate, Communication, Continuous Development, and Mission Accomplishment.

Thematic Coding

After developing NLBM 1.0, we aimed to further refine the model through a thematic coding analysis. Raw interview and focus group responses used in the qualitative card sort were analyzed with additional raw interview statements (n = 2,131) that were acquired from Navy communities. Each coder was provided with the NLBM factors and operational definitions and were given instructions, such as using a drop-down box to select which of the NLBM factors were best represented in the statement. Coders had an option to suggest that a statement should be deleted because it was not relevant to naval leadership and an option to suggest that a statement belonged to an “other” category not evident in the model.

Two researchers coded each statement. If the two researchers agreed about which leadership factor was being represented in a statement (53% of the statements), that statement was not coded a third time. But statements where

coders could not reach agreement were then coded by a third researcher. This additional coding resulted in 26% of statements reaching partial agreement (i.e., 2/3^{rds} agreement). These coding efforts resulted in 21% of all statements having no agreement across three coders. Additionally, intercoder reliability (ICR) was computed for the full set of data (Krippendorff, 1970; 2019). Krippendorff's Alpha (α) was chosen to measure ICR because of its ability to account for any number of coders, a variety of sample sizes, and nominal data such as NLBM factors (Hayes & Krippendorff, 2007; Krippendorff, 2019; Marzi et al., 2024). Using 95% confidence interval calculations from 1,000 bootstrap iterations, our findings indicated poor overall agreement (α = .51, 95% CI [.50, .52]) based on standards for satisfactory (.80) or moderate (.79 - .67) levels of agreement (Krippendorff, 2019).

Although our coder agreement percentage and ICR coefficient did not meet suggested standards, we did not find these results to be fully discouraging. When considering that researchers had 18 options for coding, there was a higher degree of difficulty than in most studies using thematic coding, which typically have fewer choices. Multiple researchers have noted that larger coding frames make coding decisions difficult (Hruschka et al., 2004) and can be problematic because coders may struggle to keep all codes in working memory (O'Connor & Joffe, 2020). Based on these results, we decided that combining the functional knowledge from the card sort analysis and NLBM 1.0 thematic coding with the right comprehensive model of leadership could lead to a better product moving forward.

Campbell's (2012b) model of leadership provided an ideal framework to refine NLBM 1.0 for a couple of reasons. First, integrating this model from research shifted the focus from a combination of individual differences and behavior to a greater focus specifically on observable behaviors, which would allow the factors to be more easily assessed and trained. Second, Campbell's model was derived from an extensive review of leadership research combined with prior longitudinal research in the Army examining the behavioral determinants of job performance across numerous specializations ("Project A"; Campbell & Knapp, 2010). This version of Campbell's (2012b) model used task performance, leadership performance, and management performance determinants, with each determinant having between six and eight factors. Combining factors from NLBM 1.0 and the model derived from Campbell's work (2012b) resulted in the NLBM 2.0. The NLBM 2.0 was organized using three meta-factors based on Campbell's (2012b) performance determinants (e.g., mission, leadership, and management), with each determinant having seven factors.

Another round of thematic coding was used to determine the effectiveness of the NLBM 2.0. Similar to the first round of coding, researchers again indicated which factor was best represented in each raw data statement (i.e., primary factor). However, this time researchers had the option to provide a second factor if they thought multiple factors were evident within a single statement (i.e., secondary factor). One group of researchers used NLBM 2.0 to code the 21% of statements that did not reach either partial or full agreement in the first round of coding (n = 793). Concurrently, two Navy SMEs coded most of the raw interview data statements (n = 3,481) using the original leadership determinants and factors from Campbell (2012b). The rationale for using two different coding frameworks was to determine if coder agreement percentage and ICR would be higher using the using the NLBM 2.0 or the original Campbell factors.

The researchers using NLBM 2.0 as a coding framework reached agreement on the primary leadership factor being represented for 40% of statements and were able to find some level of agreement (e.g., primary/secondary factor agreement, secondary factor agreement) for 64% of statements. The Navy SME coders agreed on the primary leadership factor being represented for 48% of statements and found some level of agreement for 61% of statements. ICR was then calculated by recoding both primary and secondary agreement as overall agreement. For NLBM 2.0 coding, ICR remained below acceptable levels of agreement (α = .57, 95% CI [.54, .61]). Navy SME coders had slightly improved ICR yet remained below acceptable levels of coder agreement (α = .63, 95% CI [.61, .64]). At face value, the Campbell factors coded by the Navy SMEs appear to be better supported than the NLBM 2.0 factors. However, the Campbell factors (48% primary agreement) had slightly less agreement than NLBM 1.0 factors (53% agreement). Meanwhile, the NLBM 2.0 factors demonstrated utility when applied to the most ambiguous statements in the data set, which had no agreement in prior coding efforts (i.e., NLBM 1.0). These results, combined with an adjustment to Navy needs and culture, led to the NLBM 2.0 moving forward for further refinement.

During another two-day workshop led by the Office of Naval Research (ONR), SMEs and Navy stakeholders reviewed the NLBM 2.0. The research team collected feedback on the breadth and depth of content, inclusion or exclusion of model items, factor titles and definitions, and fit with the Navy. This feedback was used in conjunction with natural language processing analysis (topic modeling) to create NLBM 3.0.

Topic Modeling

We applied a natural language processing method (NLP) as a supplement to the iterative card sort and thematic analyses because multi-method approaches to research (e.g., human-driven analyses and NLP methods) can bolster the trustworthiness of results when they converge (Rodriguez & Storer, 2020). NLP methods provide another way of analyzing the raw NLBM leadership statements that have unique advantages (e.g., decision rules are explicit) and disadvantages (e.g., lack of nuance and understanding). Ultimately, NLP methods offer an alternative means of describing the data that is less reliant on the subjectivity of human judgments.

Structural topic models are a type of NLP that can make inferences about the underlying (i.e., latent, not-directly-observable) topic structure of text. These models describe each statement as a mixture of topics and each topic as a mixture of words. We used the STM package in R for all topic models (Roberts et al., 2019). Perhaps the most important choice when creating a topic model is selecting the number of topics (k) because k affects the way a topic model categorizes data and must be set in advance. However, there is no "correct" answer to the question of how many topics are in a set of text. There are a variety of methods for inferring a reasonable range for the number of topics (Arun et al., 2010; Cao et al., 2009; Deveaud et al., 2014; Griffiths et al., 2004). We primarily focused on conducting a parallel model search (i.e., running a series of models with different k values) and comparing variables like topic exclusivity (how unique a topic is) and semantic coherence (how often related words are used together, which is correlated with human judgments of topic quality [Mimno et al., 2011]) as a guide. Ultimately, k values should be chosen based on the interpretability of the generated topics and the theory-driven considerations of the researchers. After evaluating a series of different models (e.g., k = 5, k = 21, k = 30) and diagnostics, we concluded that a 15-topic model was an appropriate choice for the NLBM statements because it maximized the interpretability of the generated topics.

Words from the topic model results that were highly frequent in a topic but also relatively rare in the text (exclusive) were chosen as distinguishing terms. Notably, researchers were quickly able to generate topic labels from the k = 15 distinguishing terms that were subsequently confirmed by reviewing the raw statements most strongly associated with each topic by the model. In addition, researchers compared the topics generated by the structural topic model to the human-derived leadership factors from NLBM 1.0 and 2.0. The topic model results supported all but one of the 18 NLBM 1.0 categories and all but four of the 21 NLBM 2.0 categories. These findings further support the structure of the NLBM model iterations by demonstrating that human and algorithmic decision makers are coming to very similar conclusions about the topics that are important for effective Navy leadership and aligned with the research literature.

Refinements leading to NLBM 3.0

The second workshop with Navy SMEs along with the findings from the thematic coding analyses provided insights that led to the development of the current model: NLBM 3.0. At the workshop, small groups worked to refine and clarify NLBM determinants, factors, and operational definitions, offering suggestions for how to make titles and descriptions more applicable to the Navy. In addition, Navy SMEs provided ideas for assessments and training interventions to align with the NLBM. Combining these ideas for improvement with results from thematic coding and the topic modeling process led to a more streamlined and effective NLBM 3.0, which consists of 22 factors across four determinants. The model with operational definitions is outlined in Tables 1a-1d.

Another model refinement from NLBM 2.0 to NLBM 3.0 that was identified via the workshop was the addition of a fourth determinant, *Ethos*, focused on measurable internal individual differences vice the observable behavior focus of the other three. The Ethos determinant contains seven factors that outline internal attributes and variables that vary from one individual to another but are relatively stable, reliably predictable, measurable, and quantifiable. Though NLBM 1.0 included individual differences, these were scattered throughout the factors in combination with observable behaviors. NLBM 3.0 reinstates these important leadership characteristics but disentangles internal traits from observable behavior.

A major goal of NLBM is to enhance Navy talent management efforts. Subsequently, the Ethos factor outlines individual difference characteristics that influence observable behavior and can be assessed through commonly used metrics (e.g., personality inventories, cognitive ability tests). The addition of Ethos with the three other behaviorally-focused factors will help the Navy align appropriate assessments to each factor, select the right Sailors for the right leadership roles by linking measurable individual traits with observable behaviors and critical

outcomes, and match development recommendations to trainable characteristics. Tables 1a-d provide an overview of the NLBM 3.0 determinants, factors, and operational definitions.

Table 1a. NLBM 3.0: Mission Determinant

Leadership Determinant #1: Mission *Observable actions directed toward the achievement of substantive Navy goals (not content-free goals such as "improve relationships").*	
Factors (5)	Operational Definition
Technical Proficiency	The degree to which the Sailor performs the core substantive technical tasks central to their role. Demonstrates technical expertise (e.g., SMEs) within and/or across teams to meet mission objectives.
Communication	The proficiency with which the Sailor conveys understandable, well-organized messages to others, demonstrates active listening skills to receive messages from others, and uses feedback to ensure meaning or intent is understood.
Initiative, Persistence, & Effort	The consistency of a Sailor's effort day-to-day, the frequency with which the Sailor expends extra time and energy required to achieve the mission, and the willingness to continue working under adverse conditions. The degree to which the Sailor is resilient, committed, and motivated to accomplish stated and implied mission objectives.
Facilitating Team & Peer Performance	The degree to which the Sailor supports peers with job-related issues and helps the team and command achieve objectives. The proficiency with which he or she builds and sustains a command that is respectful, collaborative, inclusive, diverse, and fair.
Serves as a Role Model	Upholds Navy policies and ethical standards. Demonstrates commitment to orders and directives while also providing constructive criticism to improve the organization. The degree to which the Sailor develops oneself to improve knowledge, performance, and reduce limitations.

Table 1b. NLBM 3.0: Develop/Inspire Determinant

Leadership Determinant #2: Develop/Inspire *Direct interpersonal actions taken to influence behavior of Sailors such that their performance is enhanced, both individually and collectively.*	
Factors (5)	Operational Definition
Sailor Support	The degree to which the Sailor provides recognition and encouragement. Balances mission accomplishment priorities with the personal well-being and family needs of Sailors in their command.
Relationship Building	The degree to which the Sailor builds collaborative networks and teams inside and outside of their command as well as up and down the chain of command. Fosters a positive reputation with Navy stakeholders (e.g., civilian workforce, other government agencies, non-government organizations, the "public").
Goal Emphasis	Conveys enthusiasm and commitment toward accomplishing mission objectives. Inspires others and encourages sense of purpose, meaning, and pride in the mission.
Empowerment	Delegates authority and responsibilities to others. Encourages participation. Allows discretion in decision making.
Training, Coaching, & Mentoring	The degree to which the Sailor coaches others regarding job-relevant tasks (technical, interpersonal, etc.). Offers one-on-one instruction and constructive feedback designed to improve Sailor performance. Supports leveraging mistakes as opportunities for learning and growth. Provides information and guidance about career advancement. Shares experiences and knowledge.

Table 1c. NLBM 3.0: Management Determinant

Leadership Determinant #3: Management *Observable actions that best guide and manage the use of the Navy's resources to achieve its goals. They can involve cognitive and communicative processes that influence others, but they do not rely on interpersonal influence.*	
Factors (5)	Operational Definition
Decision Making & Problem Solving	Makes timely and optimal decisions. Gathers information, assesses risk, and identifies priorities. Adjusts to unpredictable and high-pressure situations. Is open-minded and seeks innovative solutions.
Strategic Goal Setting	Sets strategic, tactical, and operational goals to accomplish current mission objectives. Formulates plans to accomplish goals. Forecasts trends and develops vision for addressing future goals.
Provides Direction, Guidance, & Structure	Coordinates and clarifies task assignments, roles, and methods within and across units. Allocates and organizes manpower, logistics, technical support, knowledge sharing, and resources. Schedules tasks and/or operations.
Monitoring Unit Effectiveness	Evaluates progress toward individual, team, and command goals and objectives. Holds self and others accountable for meeting standards. Monitors costs and resource consumption.
Administration	Performs day-to-day administrative tasks and makes information available in a timely manner. Maintains accurate records, documents actions, organizes and analyzes information.

Table 1d. NLBM 3.0: Ethos Determinant

Leadership Determinant #4: Ethos *Attributes and variables that vary in expression from one individual to another but are relatively stable, reliably predictable, measurable, and quantifiable (e.g., "Big 5" traits of personality).*	
Factors	Operational Definition
Emotional Intelligence	Demonstrates awareness and control of their own emotions and the impact their reactions have on others. Is attuned to Sailors' emotional responses and offers encouragement when needed.
Personality	The unique pattern of characteristic thoughts, feelings, and behaviors that distinguish one individual from another. It primarily encompasses traits, attitudes, and other enduring aspects of an individual's psychological makeup, which influence behavior and interactions with the environment.
Cognitive Ability	A person's capacity to perceive, process, understand, and apply information. It encompasses various mental abilities such as reasoning, problem-solving, memory, attention, and linguistic skills.
Values	Individual characteristics and fundamental beliefs that serve as a behavioral guide for an individuals' attitudes, actions, and decisions. They influence one's thought processes, priorities, and goals. Examples include characteristics such as honesty, integrity, fairness, autonomy, and social responsibility.
Interests	Individuals' preferences, inclinations, and attractions toward specific activities, topics, or domains. They reflect personal likes, passions, and motivation. Examples include areas such as technology, arts, science, leadership, helping others, or outdoor activities.
Metacognition	Awareness and understanding of one's own cognitive and emotional processes, including knowledge about how to plan, monitor, and regulate one's thinking and learning. It involves the ability to reflect on one's own mental strategies, reflect on one's knowledge and capabilities, assess one's strengths and weaknesses in problem-solving or decision-making tasks, and adapt one's approach to achieve better outcomes.
Motivation	Internal processes that drive and sustain individuals' behavior toward achieving specific goals or outcomes. It encompasses the factors that energize, direct, and maintain behavior, including needs, desires, goals, and incentives. Motivation can be intrinsic, arising from within the individual (e.g., personal satisfaction, sense of accomplishment), or extrinsic, influenced by external rewards or consequences (e.g., salary, recognition).

DISCUSSION

The NLBM, a customized and data-driven approach to leadership modeling, was built from a foundation in the research literature, developed through Navy SME input, and refined and validated using a diverse set of methods and procedures, including card sorting, thematic coding, and algorithm-based topic modeling. This multi-faceted approach provides an empirically-based framework for a long-term leadership development, training, and selection solution.

Because data is used from across naval communities, the U.S. Navy can replace the current siloed approach with a leadership model that will evaluate leaders on a set of Navy-wide factors while enabling Navy communities to emphasize the assessment and development of specific factors based on their needs. The collaboration across communities and the validation of the model factors using multiple modalities helps ensure a model that fits Navy Sailors and allows for the assessment and future development of Navy leaders.

The findings highlight several limitations to this research that deserve further examination. One such area is the distribution of agreed upon raw interview statements among NLBM factors. Some specific factors require additional discussion for their high, low, or non-existent frequency. A few factors stood out for their high frequency of responses, including Serves as a Model (689), Communication (428), and Emotional Intelligence (410), each of which contained greater than 10% of all agreed upon statements. While the prevalence of these factors could be attributed to their importance to effective leadership, another possibility is that these factors contain multiple factors. Additional data from Navy communities may resolve this concern and provide clarity.

Conversely, during NLBM 2.0 coding a few factors had low frequencies (e.g., Administration, Self-Development) and in some cases no support (e.g., Non-Job-Specific Task Proficiency, External Representation). Simply removing factors for not being coded often enough seems straightforward. However, that perspective leaves out some important nuance, including the possibility that interview guides or topical discussions simply did not highlight these factors. Instead, asking if a particular behavior or attribute is necessary to being an effective Navy leader may be more productive. However, asking for this feedback may activate implicit theories of leadership (i.e., Sailors' perceptions and expectations of leaders). These perceptions have been found to have a consistent and significant influence on perceived leader effectiveness that is often skewed toward leaders with perceived masculinity and charisma (Offermann & Coats, 2018). Looking at it from this perspective, there are certainly persuasive arguments to retain the aforementioned factors with lower frequencies. Further, each of these factors were derived from the extensive body of leadership literature (Campbell, 2012b). In short, weighing the balance between the efficiency (i.e., not having too many concepts), model extensiveness (i.e., not missing key leadership aspects), and relevance to the Navy will be vital to molding the NLBM as it continues to evolve.

Topic modeling within NLP analyses also indicated support for the NLBM factors, with slightly more support for NLBM 1.0. This support could be attributed to NLBM 1.0 being formulated from raw interview data, whereas NLBM 2.0 framework comes more from research (Campbell, 2012b). Still, NLBM 2.0 was also supported in the NLP analysis. This may indicate that the melding of the Campbell model factors with NLBM 1.0 factors to create NLBM 2.0 may have struck a successful balance between the raw interview responses and the large body of knowledge on leadership. Additional support for this conclusion also can be found in NLBM 2.0 coding, where 40% agreement was achieved for raw interview statements that three coders could not previously reach agreement upon.

The refinement to NLBM 3.0 combined NLP analysis with SME-input and the research foundation. This appears to provide an excellent balance of leadership theory, implicit perceptions, and Navy needs. As more data is collected, further refinement will offer further insight into the validity and utility of the NLBM as it evolves.

CONCLUSION

This paper adds to the field of leadership research in two key ways. First, the study demonstrates a rigorous method for developing and validating behavioral models of leadership that goes beyond theoretical findings. Reliance on theory alone to develop a behavioral leadership model runs a high risk of failure, as no one theory can account for the constellation of phenomena that impacts leadership (Rumsey, 2012). Instead, a combination of choosing the best theoretical model from the vast array of options along with analyzing data from a broad range of relevant organizational sources can provide stronger and more relevant, applicable solutions for the specific organization than relying on theory alone. This is particularly true when the solution is tied to leader assessment, training, and selection

within an organization. The NLBM meets these criteria by being developed with a relevant theoretical model, using qualitative data from Navy personnel about their leadership experiences within naval communities, and through ensuring that leadership factors are measurable and trainable to align with Navy talent management goals.

Second, the study afforded the opportunity to incorporate NLP algorithms and human analyses on the same large set of data. While the NLP results helped confirm the findings of the human analysis at the factor level, they lacked the richness and nuance of human analysis in being able describe how each factor added to the model as a whole. Human judgments about model efficacy and topic content are still needed. The human analysis, on the other hand, was labor intensive, subjective, and reaching sufficiently high levels of inter-coder agreement was difficult due to the noise in the data and the model's complexity. Notably, both approaches depend on the accuracy and clarity of the data, which reinforces the need to maximize measurement accuracy in data collection as much as possible. A more optimal path might be to start with NLP analyses and use the results as a baseline to drive theoretical model development and subsequently apply more focused human-analyses (Rodriguez & Storer, 2020). Such an approach would likely build on the strengths and bolster the weaknesses of both approaches.

As the work continues, we will continue to leverage the combination of statistical analysis such as NLP with SME review to derive a validated and meaningful leadership model. In addition to further refinement with new data into NLBM 4.0, we hope to conduct criterion validity testing to evaluate and determine its relevance to important Navy outcomes and empirically justify its framework.

REFERENCES

Alldredge, M. E., & Nilan, K. J. (2000). 3M's leadership competency model: An internally developed solution. *Human Resource Management, 39*(2–3), 133–145. https://doi.org/10.1002/1099-050X(200022/23)39:2/3<133::AID-HRM4>3.0.CO;2-8

Arun, R., Suresh, V., Veni Madhaven, C. E., & Narasimha Murthy, M. N. (2010). On finding the natural number of topics with latent dirichlet allocation: Some observations. In. M.J. Zaki, J. X. Yu, B. Ravindran, & V. Pudi (Eds.), *Advances in Knowledge Discovery and Data Mining*. PAKDD 2010. Lecture notes in Computer Science, 16. Springer. https://doi.org/10.1007/978-3-642-13657-3_43

Avolio, B., & Bass, B. (1991). *The full range of leadership development: Basic and advanced manuals*. Bass, Avolio, & Associates

Campbell, J. P. (2012a). Leadership, the old, the new, and the timeless: A commentary. In M. G. Rumsey (Ed.), *The Oxford handbook of leadership* (pp. 401–419). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195398793.013.0024

Campbell, J. P. (2012b). Behavior, performance, and effectiveness in the twenty-first century. In S. W. J. Kozlowski (Ed.), *The Oxford handbook of organizational psychology,* (pp. 159-196). Oxford University Press.

Campbell, J. P., & Knapp D. J. (2010). Project A: 12 years of R&D. In J.L. Farr & T. Tippins (Eds*.), Handbook of employee selection* (pp. 865-886). Routledge.

Cao, J., Kia, T., Li, J., Zhang, Y., & Tang, S. (2009). A density-based method for adaptive LDA model selection. *Neurocomputing, 72*(7-9), 1775-1781. https://doi.org/10.1016/j.neucom.2008.06.011

Cigularov, K., & Navy Expeditionary Combat Command (2021). *EOD leader behavioral competency model interview notes* (Unpublished raw data).

Department of the Air Force. (2022, February). *Total force development and management: Air Force Handbook [AFH] 36-2647*. https://static.e-publishing.af.mil/production/1/af_a4/publication/afi36-2650/afi36-2650.pdf

Department of the Army. (2022, November). *Field Manual 6-22: Developing Leaders*. https://armypubs.army.mil/epubs/DR_pubs/DR_a/ARN36735-FM_6-22-000-WEB-1.pdf

Department of the Navy. (2019). *Navy Leader Development Framework: Version 3.0* https://media.defense.gov/2020/May/18/2002302036/-1/-1/1/NLDF3MAY19.PDF

Department of the Navy. (2020). *Signature behaviors of the 21st Century Sailor: Version 2.0*. From https://www.med.navy.mil/Portals/62/Documents/BUMED/SG/signature-behaviors.pdf?ver=ZobKa9grGZscP7404QZBfg%3D%3D

Department of the Navy. (2021a). *Chief of Naval Operations Navigation Plan 2021.* https://media.defense.gov/2021/Jan/11/2002562551/-1/-1/1/CNO%20NAVPLAN%202021%20-%20FINAL.PDF

Department of the Navy. (2022a). *Chief of Naval Operations Navigation Plan 2022*. https://media.defense.gov/2022/Jul/26/2003042389/-1/-1/1/NAVIGATION%20PLAN%202022_SIGNED.PDF

Deveaud, R., SanJuan, E., & Bellot, P. (2014). Accurate and effective latent concept modeling for ad hoc information retrieval. *Revue des Sciences et Technologies de l'Information, 17*, 61-84. https://doi.org/10.3166/DN.17.1.61-84

Griffiths, T. L., Steyvers, M., Blei, D. M., & Tenenbaum, J. B. (2004). Integrating topics and syntax. In L. K. Saul, Y. Weiss, & L. Bottou (Eds.), *Advances in Neural Information Processing Systems* (pp. 537-544). MIT Press.

Hruschka, D. J., Schwartz, D., St John, D. C., Picone-Decaro, E., Jenkins, R. A., & Carey, J. W. (2004). Reliability in coding open-ended data: Lessons learned from HIV behavioral research. *Field Methods, 16*, 307–331. https://doi.org/10.1177/1525822X04266540

Krippendorff, K. (1970). Estimating the reliability, systematic error and random error of interval data. *Educational and Psychological Measurement, 30*, 61–70. https://doi.org/10.1177/001316447003000105

Krippendorff, K. (2019). *Content analysis: An introduction to its methodology* (4th Ed.). Sage Publications. https://doi.org/10.4135/9781071878781

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures, 1,* 77–89. https://doi.org/10.1080/19312450709336664

Marzi, G., Balzano, M., & Marchiori, D. (2024). K-Alpha Calculator—Krippendorff's Alpha Calculator: A user-friendly tool for computing Krippendorff's alpha inter-rater reliability coefficient. *MethodsX, 12*, 102545. https://doi.org/10.1016/j.mex.2023.102545

McLean, S., & Moss, G. (2003). They're happy, but did they make a difference? Applying Kirkpatrick's Framework to the evaluation of a National Leadership Program. *The Canadian Journal of Program Evaluation*, *18*(1), 1–23. https://doi.org/10.3138/cjpe.018.001

Mimno, D., Wallach, H. M., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. *Proceedings for the 2011 Conference on Empirical methods in Natural Language Processing*, 262-272.

Naval Submarine Medical Research Laboratory. (n.d.-a). *Commander Naval Air Force job analysis traits/competencies* (Unpublished raw data).

Naval Submarine Medical Research Laboratory. (n.d.-b). *Human Resources job analysis traits/competencies* (Unpublished raw data).

Naval Submarine Medical Research Laboratory. (n.d.-c*). Submarine Command leadership job analysis traits/competencies* (Unpublished raw data).

Naval Surface Warfare Center Dahlgren. (n.d.-a). *Engineering Duty Officer Command job analysis traits/competencies.* (Unpublished raw data).

Naval Surface Warfare Center Dahlgren. (n.d.-b). *Commander Surface Forces job analysis traits/competencies* (Unpublished raw data).

Naval Surface Warfare Center Dahlgren. (2024). *Commander Surface Forces competency model interview transcripts* (Unpublished raw data).

O'Connor, C., & Joffe, H. (2020). Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods, 19*, 1-13. https://doi.org/10.1177/1609406919899220

Offermann, L. R., & Coats, M. R. (2018). Implicit theories of leadership: Stability and change over two decades. *The Leadership Quarterly, 29*(4), 513-522. https://doi.org/10.1016/j.leaqua.2017.12.003

Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). Stm: An r package for structural topic models. *Journal of Statistical Software, 91*(2), 1-40. https://doi.org/10.18637/jss.v091.i02

Rodriguez, M. Y., & Storer, H. (2020). A computational social science perspective on qualitative data exploration: Using topic models for the descriptive analysis of social media data. *Journal of Technology in Human Services*, *38*(1), 54–86. https://doi.org/10.1080/15228835.2019.1616350

Rumsey, M.G. (2012). The elusive science of leadership. In M.G. Rumsey (Ed.), *The Oxford Handbook of Leadership* (pp. 455-466). Oxford University Press, Inc.

Yukl, G., Gordon, A., & Taber, T. (2002). A hierarchical taxonomy of leadership behavior: Integrating a half century of behavior research. *Journal of Leadership & Organizational Studies, 9*(1), 15–32. https://doi.org/10.1177/107179190200900102