Faculty and Researchers | Faculty and Researchers' Publications

2022

Performance Evaluation Trait Validation

Helzer, Erik; Bacolod, Marigee

Monterey, California: Naval Postgraduate School

https://hdl.handle.net/10945/71855

NAVAL RESEARCH PROGRAM
NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

PERFORMANCE EVALUATION TRAIT VALIDATION EXECUTIVE SUMMARY

Principal Investigator (PI): Dr. Erik G. Helzer, Department of Defense Management (DDM)

Additional Researcher(s): Dr. Marigee Bacolod, DDM

Student Participation: LTJG Bryan Luke, USN, DDM; LT Phillip Gervato, USNR, DDM

Prepared for:
Topic Sponsor Lead Organization: N1 - Manpower, Personnel, Training & Education
Topic Sponsor Organization(s): Navy Personnel Command
Topic Sponsor Name(s): CAPT Ben Baran, PhD, SPHR, Performance Evaluation Transformation Coordinator, Navy Personnel Command
Topic Sponsor Contact Information: benjamin.e.baran.mil@us.navy.mil

Project Summary

We present the results of a year-long study supporting the Navy's efforts to transform its existing Performance Evaluation System (PES) with an evaluation tool that has been updated to reflect current Navy values and priorities. Our team validated and refined trait and value statements (TVS) developed for rating Sailors on a variety of dimensions, including leadership, teamwork, communication, resilience, and other character strengths. In addition, our team developed rating scales for assessing Sailors' future potential for performance at the next paygrade, as well as a comparative assessment tool for evaluating Sailors against their peers. We then conducted a large-scale prototyping study, which involved the development of a performance evaluation prototype instrument, recruitment of Sailors-reported-on and performance raters, and statistical analysis of the properties of the prototype instrument—in particular, the measurement validity of the TVS. Using data from 606 performance raters, we find that ratings of Sailors based on past performance and future potential predict comparative assessments, with some traits (e.g., leadership) better distinguishing the top from middle performers. Few of the trait ratings distinguished among bottom performers, although such performers were relatively rare in our sample. Ratings based on past performance and future potential were correlated with one another; analyses indicate they convey unique information about the Sailor-reported-on, but that both predicted workplace behaviors. Based on our research, we offer four concrete recommendations for action. First, we recommend amplifying the comparative assessment in future PES. Second, the Navy should consider adopting ratings of future potential for developmental/coaching purposes using actionable, concrete developmental feedback. Third, future evaluation instruments should focus on a subset of performance and/or future potential traits in assessing job performance to keep PES simple and useful. Finally, we recommend assessing the predictive validity of these new measures using multi-source data.

Keywords: *performance evaluation system, performance assessment, traits, job performance, measurement validity, reporting senior's cumulative average, RSCA*

Background

A valid and credible PES is critical for identifying and managing talent in the U.S. Navy. The Navy's *Sailor 2025* initiative called for an updating of the Navy's personnel management system to reflect the Navy's current goals with respect to recruitment, retention, and advancement (Burke, 2018). In addition, concerns expressed by Task Force One Navy point to the need for a PES that is both fair and objective (Task Force One Navy, 2021). Major efforts to generate updated performance trait statements that are consistent with Navy doctrine, instructions on performance appraisal, and Sailor values were conducted in 2002 (under the Task Force Excel 5-Vector Model) and 2019 (by researchers associated with the performance evaluation transformation initiative). In 2019, a working group identified TVS to serve as the basis of revised performance evaluation metrics. This set of TVS modernizes the performance criteria against which Sailors would potentially be evaluated; however, a systematic study on the validity of this set of TVS has not been conducted.

We executed a year-long study supporting the Navy's efforts to transform the existing PES with an evaluation tool that has been updated to reflect current Navy values and priorities. We validated and refined the TVS, benchmarking items against current Navy doctrine and performance evaluation materials from other military services. We then conducted a large-scale prototyping study, which involved the development of a performance evaluation prototype instrument,

recruitment of Sailors-reported-on and performance raters, and statistical analysis of the properties of the prototype instrument—in particular, the measurement validity of the TVS.

We recruited Sailors-reported-on through two different channels, including a large-scale social media advertisement. These Sailors (N = 1823) furnished the names and contact information of performance raters (e.g., supervisors, peers) who could meaningfully evaluate Sailors' job performance. Performance raters were directed to an online platform allowing them to evaluate Sailors using the performance evaluation prototype. Usable responses were collected from 606 performance raters and were statistically analyzed to assess the validity of items comprising the performance evaluation prototype.

Findings and Conclusions
Ratings of traits related to past performance and future potential were all correlated with measures of positive and negative workplace behaviors in the expected directions. Those rated highly on performance and potential tended to engage in organizational citizenship behaviors more frequently and counterproductive workplace behaviors less frequently than their peers.

We assessed whether the trait ratings had predictive validity using survey responses on the comparative assessment as the criterion. All the past performance traits had some predictive power, and some traits were better at distinguishing the top from the middle of the performance distribution. For example, ratings of *leadership* and *initiative* based on past performance were better able to distinguish the top from the middle performers, but ratings of *character* and *resilience* based on past performance were better at distinguishing the bottom from the middle performers. On the other hand, the future potential trait of *experience and competence* was able to distinguish both top and bottom performance.

A. Recommendation 1: *Amplify the comparative assessment in future PES.*

To differentiate talent and provide useful information to promotion boards and talent managers, we recommend amplifying the comparative assessment in future PES. Of all the job performance indicators assessed in our prototype, none was as informative in differentiating performance as the comparative assessment.

B. Recommendation 2: *Consider adopting future potential ratings for developmental/coaching purposes, especially the word blocks for actionable, concrete developmental feedback.*

Trait items based on future potential provide a unique source of information that may be ideal for the purpose of development and coaching. In this study, Sailors were evaluated on six trait dimensions (e.g., *character development*, *leadership and teamwork skills*, *judgment and decision-making*) based on their readiness to succeed at the next paygrade or in key, particularly demanding jobs. The benefit of this rating system is that it is forward-focused and constructive. When Sailors were rated as "not ready," we prompted respondents to offer concrete, actionable feedback to help the Sailor move toward greater perceived readiness. These trait ratings and qualitative feedback could serve as the basis for coaching sessions with Sailors to identify focus areas for job performance improvements that can advance them toward their next career milestone.

C. Recommendation 3: *Focus on a subset of performance and/or potential traits in assessing job performance to keep PES simple and useful.*

To reach the goal of transforming the PES toward a model that is simple, useful, and fair, we recommend focusing on a subset of the traits rating past performance and/or those rating future potential. Ratings of past performance are potentially simpler and more useful indicators of job performance than ratings of future potential.

Recommendations for Further Research
Our current efforts are focused on merging data from Sailors' personnel records with data from our prototype testing to expand our view of predictive, convergent, and divergent validity of the trait and value statements (TVS). These data and resulting analyses are critical to understanding how well ratings of Sailors' job performance using the refined set of TVS reflect and track with job performance on a longer time horizon across multiple indicators of performance. These data will also allow us to carefully examine potential adverse impacts of the refined TVS, including the question of whether some TVS inappropriately favor certain groups over others. We caution against implementing the TVS into performance evaluation tools until the results of these analyses are obtained.

References

Burke, R. P. (2018, March). Sailor 2025: The Navy's strategy for people. *Proceedings of the U.S. Naval Institute*. https://www.usni.org/magazines/proceedings/2018/march/sailor-2025-navys-strategy-people

Task Force One Navy. (2021). *Task Force One Navy: Final Report*. https://media.defense.gov/2021/Jan/26/2002570959/-1/-1/1/TASK%20FORCE%20ONE%20NAVY%20FINAL%20REPORT.PDF

Acronyms
PES   Performance Evaluation System
RSCA   reporting senior's cumulative average
TVS   trait and value statements