Performance Evaluation and Promotion Recommendation: Best Practices for Selection Boards Phase II

NPS-DDM-26-001

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

PERFORMANCE EVALUATION AND PROMOTION RECOMMENDATION: BEST PRACTICES FOR SELECTION BOARDS, PHASE II

by

Dr. Marigee Bacolod and Dr. Latika Hartmann

February 2025

Prepared for: N1 Manpower, Personnel, Training & Education. This research is supported by funding from the Naval Postgraduate School, Naval Research Program (PE 0605853N/2098). NRP Project ID: NPS-25-N061-A

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED	
February 1, 2026	Technical Report	START DATE 10/1/2024	END DATE 3/31/2026

4. TITLE AND SUBTITLE
Performance Evaluation and Promotion Recommendation: Best Practices for Selection Boards, Phase II

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER
		0605853N/2098
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**
NPS-25-N061-A		

6. AUTHOR(S)
Dr. Marigee Bacolod and Dr. Latika Hartmann

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Naval Postgraduate School 1 University Circle Monterey, CA 93943-5000	NPS-DDM-26-001

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)	11. SPONSOR/MONITOR'S REPORT NUMBER(S)
N1 Manpower, Personnel, Training & Education Naval Postgraduate School, Naval Research Program	N1; NRP	NPS-25-N061-A NPS-DDM-26-001

12. DISTRIBUTION/AVAILABILITY STATEMENT

Distribution Statement A. Approved for public release: Distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

In collaboration with the Navy Talent Management Center of Excellence, we developed and tested two prototype displays of a fictional Performance Summary Record (PSR) that can be used at Selection Boards. We randomly assigned two groups of NPS students to view each display, answer a knowledge check on the candidate evaluation shown on the display, and finally complete a user experience survey. We found few significant differences in responses to the knowledge check between the two groups. While NPS participants assigned to the Horizontal prototype were more favorable to their display, both groups assigned near-identical promotion scores to the fictional candidate. Testing of subjects at Navy Personnel Command at Millington also found similar results—few significant differences in knowledge check across the two groups. However, the Millington subjects were less positively inclined to the Horizontal display. Since a large share of the Horizontal group at Millington was higher-ranked Navy officers with extensive Board experience, their experience likely biased their opinions of the different layout of the PSR in the Horizontal display. Participants in both locations identified visual cues as useful design features along with the specific performance metrics of comparative assessment and relative value. As the Navy continues to refine and test future prototypes, we recommend incorporating the visual cues into future PSRs to reduce the cognitive burden for Selection Board members.

15. SUBJECT TERMS
Selection Boards, Navy Performance Evaluation, displays

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
a. REPORT	b. ABSTRACT	C. THIS PAGE		
U	U	U	UU	61

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER *(Include area code)*
Marigee Bacolod	831-656-3302

STANDARD FORM 298 (REV. 5/2020)
Prescribed by ANSI Std. Z39.18

THIS PAGE INTENTIONALLY LEFT BLANK

NAVAL POSTGRADUATE SCHOOL
Monterey, California 93943-5000

Ann E. Rondeau
President

Matt Carlyle
Acting Provost

The report entitled "Performance Evaluation and Promotion Recommendation: Best Practices for Selection Boards, Phase II" was prepared for N1-Manpower, Personnel, Training & Education and funded by the Naval Postgraduate School, Naval Research Program (PE 0605853N/2098).

This report was prepared by:

Marigee Bacolod
Associate Professor

Latika Hartmann
Associate Professor

Reviewed by:

Simona Tick, Acting Chair
Department of Finance, Acquisition and Manpower

Released by:

Michael Hesse,
Vice Provost for Research & Innovation

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. BACKGROUND

U.S. Navy promotion and Selection Boards utilize performance evaluation reports as primary inputs for determining whom to promote among the best and most fully qualified talent in the Navy. These reports contain a large amount of information presented in a form that does not readily facilitate rapid, accurate information processing. In FY24, Bacolod and Helzer (2025) conducted Phase I of this project, wherein they sought to identify how Selection Board members process information presented to them in the Official Military Personnel File (OMPF), Performance Evaluation Summary Records (PSR), and information displays. In particular, the team documented the pain points in processing information that potentially lead to cognitive errors and biases in decision-making under cognitive load. The outcome of Phase I identified time constraints and lay judgment (personal standards) that had the potential to introduce cognitive noise into the Selection Board process.

B. RESESARCH OBJECTIVES AND METHODOLOGY

Following Phase I, our Phase II project goal was to further develop and test prototype displays of PSRs that can be used in Navy Selection Boards incorporating the findings of Phase I. This work builds on past efforts between Naval Postgraduate School (NPS) and the Navy Performance Evaluation and Transformation team to support ongoing changes to the Performance Evaluation System (Ahn and Hartmann 2022, 2023, 2025; Bacolod and Helzer 2022, 2023, 2025; Corbett 2024 and Hill 2025). We direct interested readers to those reports and theses to learn more about the Navy Performance

Evaluation System, Navy promotion and Selection Boards, and determinants of performance and promotion.

Our primary research questions for this project are as follows:

1. What features of the decision/information environment improve Selection Board members' ability to scan, absorb, understand, interpret, and decide in the Tank voting process?

2. What are best practices for implementing the insights above into Navy Selection Board processes?

3. How do we equate the data elements and visual displays between existing forms and the potential future state of performance evaluation?

To answer these questions, we worked with the team at the Navy Talent Management Center of Excellence to develop two prototype displays that could be used at Selection Boards. We then tested differences in information processing and user experience across the two displays on testing subjects at NPS and Navy Personnel Command in Millington, Tennessee. To evaluate differences in information processing, we randomized the participants at NPS into two groups. One group was shown a Horizontal prototype display wherein the candidate evaluation is oriented horizontally, moving from older evaluations on the left to the most recent evaluation on the right. The other group was shown a Vertical prototype display mirroring the current display of PSRs shown in Selection Board deliberations. We then asked both groups to answer questions about the fictional candidate's background and evaluation, i.e., we conducted an accuracy and knowledge check. After answering these questions, participants completed a user experience survey to share their views on helpful and unhelpful aspects of the display.

C. FINDINGS

We found few differences in the responses to the knowledge check between the Horizontal and Vertical-Traditional groups. Apart from a couple of questions on evaluation trends, there were no significant differences between the Horizontal and Vertical groups. Moreover, participants in both groups assigned near-identical promotion scores to the fictional candidate. However, participants in the Horizontal group were more favorably inclined to the prototype display compared to those in the Vertical group. As few NPS participants had prior Navy Board experience, they were able to offer a fresh perspective unbiased by current displays used at the Boards.

Participants in both groups emphasized the need for stronger visual cues, color coding to highlight changes in performance, and larger font size on the record. These changes would reduce the cognitive burden involved in quickly processing the large amount of information shown on a PSR. Both groups also found the new proposed performance metrics for comparative assessment and relative value to be useful features of the design.

Due to Department of Defense (DoD) restrictions on travel, we were unable to conduct the testing at Millington, and the testing was led by Dr. Andrea Bizarro from the Navy Talent Management Center of Excellence (PERS-00K). There were no significant differences between the two groups at Millington in the knowledge check, while participants in the Vertical group were more favorably inclined to their display. However, we hesitate to draw strong conclusions between the two Millington groups because they were unbalanced in size and Board experience. A large share of the Horizontal group were higher-ranked Navy officers with extensive Board experience, which likely biased

their opinions on the newer horizontal layout of the PSR in the Horizontal display. That said, both groups assigned almost the same promotion scores to the fictional candidate. We observed the same pattern at NPS. The Millington participants also pointed to the color coding as a useful design feature that could be further refined in newer prototypes of the displays.

As participants in both locations appreciated the visual cues on the displays regardless of their orientation, we strongly recommend the Navy incorporate color coding and perhaps arrows to highlight changes in a candidate's performance over time. Finally, an overwhelming majority of the participants were favorable to the new metrics for comparative assessment and relative value. These results bode well for the transformation efforts as the Navy builds momentum to introduce these changes in future FITREPs/EVALs and Selection Boards.

D. ORGANIZATION OF THE REPORT

This report is organized as follows. Chapter II describes the methodology, including the prototype displays and the knowledge check and user experience surveys, along with our testing protocol. Chapter III describes the findings for both the NPS and Millington groups, and Chapter IV concludes with our recommendations. Survey instruments and testing protocols are provided in Appendixes A to D. For background and related literature on Navy promotion and Selection Boards, see Chapter II of our Phase I report (Bacolod and Helzer 2025) as well as the literature reviewed more extensively in the master's theses by Corbett (2024) and Hill (2025).

II. METHODOLOGY

We worked with the Navy Performance and Evaluation Team, in particular, Dr. Andrea Bizarro and CDR Michael Major, to develop the two prototype PSR visual displays for testing along with the knowledge check and user experience survey. We describe these inputs in Sections A and B below. Then we summarize our methodology for testing between the two prototype displays in Section C.

A. PROTOTYPE VISUAL DISPLAYS

Navy Selection Boards use displays to share candidate summary information with Board Members during deliberations. This display of the PSR typically includes demographic information, billet assignments, the names of reporting seniors, EVAL/FITREP distribution of trait scores, trait averages, reporting senior cumulative averages (RSCA), and promotion recommendations such as Promotable (P), Must Promote (MP), Early Promote (EP), and so on.

In 2024, the Performance and Evaluation team at the Navy Talent Management Center of Excellence contracted with the Cognitive Performance Group to solicit feedback from Navy Selection Board members on the information in the PSRs and to develop new data visualization tools to help the decision making process during Selection Boards. To this end, Cognitive Performance Group developed multiple alternate visual displays that could be deployed during Selection Boards. The Navy Performance and Evaluation team further updated these displays incorporating the qualitative feedback from Selection Board members summarized in the Cognitive Performance Group report (Celestin et al. 2024).

We worked with the Performance and Evaluation team on two of these visual displays to use for prototype testing. Both displays used the proposed changes to FITREPs/EVALs, in particular, the new language on comparative assessment and competencies (in lieu of traits). These changes include new promotion recommendations ranging from "Not ready" to "Exceptionally ready" (see Table 1 for updated comparative assessment promotion recommendations), updated ratings on core competencies assessed on a 1 to 5 scale (see Table 1), a new relative value (RV) ranging from 50 to 100, where higher numbers are intended to capture candidates assessed as above average by their reporting seniors (relative ranking), and reporting senior remarks, i.e., bottom line up front (BLUF). As the Navy team has been testing these proposed changes to FITREPs/EVALs in the fleet, our goal was to evaluate whether a vertical or horizontal orientation of the updated performance characteristics on PSRs used in Selection Boards may lead to changes, if any, in the interpretation of these data by Selection Board members. In addition, we wanted to assess the user experience of navigating the prototype displays.

To that end, we developed two prototype visual displays in collaboration with the Performance and Evaluation team. The first display, shown in Figure 1, is similar to the current PSR. It displays a candidate's record vertically beginning from the most recent evaluation at the top and moving down to past evaluations at the bottom. The display summarizes information from the past ten years of a candidate's evaluations. Moving from left to right, the display shows information on rank, duty station, dates of evaluation and reporting senior. Next, it summarizes the promotion status, comparative assessment, relative value, and the physical fitness assessment (PFA). This is followed by a display of

the individual core competency ratings across the six proposed new competencies, and finally the reporting senior remarks. Based on past reports, the literature on decision science, and feedback from Board Members, we color coded the comparative assessment and core competency ratings. We also used the same color coding to distinguish between high and low performers for the comparative assessment and competency ratings. This was to minimize the cognitive burden in interpreting changes in these measures over time. We termed this display the Vertical-Traditional display because the layout resembles the current PSR albeit with the proposed new performance metrics.

The second display, shown in Figure 2, involves a horizontal orientation of a candidate's record. It begins with the summary information at the top on reporting senior, rank and duty station. Similar to Figure 1, the horizontal display summarizes the past ten years of evaluations, beginning with the oldest evaluation on the left and moving to the most recent evaluation on the right. Unlike Figure 1, the rows in Figure 2 capture different aspects of the evaluation while the columns correspond to different evaluation years. Below the demographic and billet information, the display shows the promotion status followed by comparative assessment, relative value and PFA score. This information is followed by the core competency ratings and ends with the reporting senior remarks at the bottom. We termed this display the Horizontal display.

To populate candidate information in these prototype displays, we used enlisted EVALs shared by prior enlisted NPS personnel. This ensured the information on the prototype displays was as close as possible to actual information that would be displayed in a Selection Board. To mimic the feel and language of "real" FITREPs/EVALs, we assigned scores to the comparative assessment and core competency ratings to match the

reporting senior remarks. Moreover, we drew on our detailed reading of multiple EVALs for different ranks to craft the reporting senior comments in the prototype displays. Finally, we also assessed the validity of these inputs with the Performance and Evaluation team along with Navy student officers at NPS.

B. KNOWLEDGE CHECK AND USER EXPERIENCE SURVEY

In the testing phase, our main goal was to assess whether individuals differentially interpret the same information across the two displays, i.e., whether the horizontal or vertical orientation of the PSR affects the data interpretation. To that end, we constructed a fifteen-question accuracy and knowledge check. This check included four questions on the demographics of the candidate followed by one question on relative value, two questions on the comparative assessment, one question on PFA, five questions on core competency ratings and a final question on what promotion score would be assigned to this candidate record. Appendix A shows the complete accuracy and knowledge check.

Finally, we also developed a user experience survey (Appendix B) to capture views on the usefulness and relevance of the information shown on the display, and the ease of use of the display. The survey included a mix of Likert scale questions and free text responses to identify helpful and unhelpful features of the display along with recommendations for changes to improve the display.

C. TESTING

We completed the development of the two prototype visual displays, accuracy and knowledge check, and user experience survey in June 2025. We then identified two groups of testing subjects: NPS students with limited Selection Board experience and Millington (TN) subjects based at Navy Personnel Command who were likely to have

more Board experience. We recruited the NPS participants at the end of June and early July and conducted the testing on the afternoon of July 17 between 1500 and 1700.

To recruit NPS participants, we began with the roster of students enrolled in MN3070, Fundamentals of Cost-Benefit Analysis, and MN4761, Applied Manpower Analysis. These courses are offered in the Department of Defense Management and typically attract students specializing in finance, acquisition and manpower. We requested their participation via email (Appendix C) and thirty-four students agreed to participate in the testing. Of these thirty-four students, we randomly assigned seventeen participants to the Vertical-Traditional display testing group and the remaining seventeen participants to the Horizontal display testing group. While we shared information on testing logistics with both groups, we did not disclose to either group their treatment (Vertical or Horizontal).

We conducted the testing on July 17 in two classrooms next to each other. One of us conducted the testing in one classroom, while the other conducted the testing in the neighboring classroom. In both cases, we used the same protocol (Appendix D). We began by introducing ourselves and the purpose of the testing. We shared a brief video prepared by CDR Major summarizing the proposed changes from promotion recommendations to comparative assessment, from trait ratings to core competency ratings, and relative values. We then displayed the prototype (Vertical in one room and Horizontal in the neighboring room) on the classroom screen and asked students to respond to the accuracy and knowledge check quiz administered via Microsoft Forms. After completing the knowledge check, we shared a QR code for the user experience survey that NPS participants also completed on Microsoft Forms. After the participants

completed the survey, we solicited informal feedback on the prototype displays. This was intended to capture any comments that participants may not have been comfortable sharing online.

Due to DoD-wide travel restrictions, we were unable to recruit the test subjects at Millington or run the testing protocol. We shared the necessary materials (testing protocol, prototype displays, knowledge check and user experience survey) with the Performance and Evaluation team that implemented the Millington testing. Our understanding is the testing was conducted in rooms similar to Selection Board session rooms. Due to differences in room logistics between NPS and Millington, the Performance and Evaluation team made changes to the Horizontal prototype display and used different color coding on both displays (see Figures 3 and 4). Unlike NPS, the knowledge quiz and user experience surveys were administered in hard copy, with participants completing them using a pen or pencil. In another difference, the Millington testing was spread over two days, July 30 and July 31, with AM and PM sessions offered on both days. After the testing was completed, we received hard copies of the knowledge checks and surveys. On account of these differences in the testing protocol between NPS and Millington, we are cautious about comparing the results across the two sites as discussed in the next section.

III. FINDINGS

A. NPS TESTING RESULTS

As described in Section II, seventeen subjects were assigned to the Horizontal display testing group and another seventeen were assigned to the Vertical-Traditional display testing group. We randomized students who agreed to participate into treatment (display type) and we followed the same testing protocol in each classroom. The key difference, i.e., treatment, was that one group of participants saw the Horizontal display prototype and the other group saw the Vertical-Traditional display prototype.

We begin by summarizing the results of the accuracy and knowledge check across the two NPS testing groups in Table 2. The top panel summarizes the characteristics of the two groups obtained from the user experience survey. Although we randomized students into the testing groups, the Horizontal group had a significantly larger share of US Navy Officers. NPS enrolls military officers from the US Army, Air Force, Coast Guard and Marine Corps, and from a small number of allied international countries. While the Navy accounted for 41% of the Horizontal group compared to 71% for the Vertical group, there were no significant differences in rank across the two groups. In both groups, 47% of the participants were O-3s. More importantly, there was no difference in Selection Board experience between the two groups—only two participants in each group (12%) had any Selection Board experience. This is unsurprising as most NPS students are junior officers.

As the sample sizes of the two groups are small (under 20), the t-tests are conservative, i.e., they are unlikely to find statistically significant differences unless there are large differences between the groups. That said, the averages on most questions are

similar for the Horizontal and Vertical groups. Both groups took just under 10 minutes to complete the knowledge check, suggesting the displays did not differentially affect the ability of participants to process the information. Most participants answered the background and rank questions correctly with no significant differences. Participants also successfully identified the specific year when the fictional candidate, Joe Seaman, received an Exceptionally Ready comparative assessment and when he passed his physical fitness assessment. Apart from two questions, there were no significant differences in correct responses across the two groups, although the Horizontal testing group answered more questions correctly on average compared to the Vertical group.

In cases where the two groups had significant differences in responses, the questions involved comparing data across multiple core competencies or comparative assessment. For example, question 10 asked participants to identify all core competencies for which Joe Seaman received a 5. The correct response was all of them across the years shown on the prototype. Only 59% of participants in the Vertical group answered this question correctly, although they did identify the subset of core competencies with scores of 5. In contrast, 94% of the Horizontal participants answered this question correctly. In the case of question 11, only 53% of participants in the Horizontal group correctly identified the number of years that Joe Seaman received a 5 across all competencies. Among the Vertical participants, only 24% answered the question correctly.

Many participants in both groups were unable to correctly describe the trends in core competency ratings over time (questions 12-14). This may be related to the wording of the questions (many responses point to confusion on what the question was asking), or to participants being unable to interpret changes in core competency ratings over time.

While the differences were insignificant across the two groups, they suggest the Navy may want to incorporate stronger visuals to highlight changes in metrics over time.

Apart from the issues on trends, the participants in both groups correctly identified the key information about the candidate and their performance on the prototype displays. Specifically, both groups assigned a similar promotion score to Joe Seaman—an average of 85 for the Horizontal and 84 for the Vertical. This result is reassuring and suggests the type of display may be unlikely to bias Selection Board decisions in any particular direction.

Next, we summarize the results of the user experience survey in Table 3, which shows that an overwhelming majority of participants in both groups responded positively to the relevance and ease of use of the prototype displays. For example, 94% of participants in the Horizontal group and 88% in the Vertical group agreed or strongly agreed that the information shown is useful for evaluating candidates. In the Horizontal group, 88% of participants compared to 77% in the Vertical group agreed or strongly agreed that it was easy to find candidate information on the displays. Finally, 100% of participants in the Horizontal group and 94% of participants in the Vertical group were confident or very confident about grading records after reviewing the records on the display.

While more participants in the Horizontal group agreed or strongly agreed with statements on the ease of use of display, the differences between the groups were not statistically significant. In the case of one question on general satisfaction, 88% of participants in the Horizontal group said they were satisfied with the experience compared to 65% of participants in the Vertical group. But this difference is due to six

out of seventeen participants in the vertical group selecting the neutral response (Figure 5, Panel E). We show the complete distribution of responses to the user experience questions in Figure 5. These graphs corroborate the summary measures in Table 3. Only one or two participants disagreed with statements covering the different aspects of user experience across the groups.

In response to the free text questions, 53% of participants in the Horizontal group said they had no difficulty locating information on the display, and others said it just took a minute to orient themselves to the information shown on the display. In contrast, only 18% of participants in the Vertical group said they had no difficulty. Most of these difficulties were on account of the small font size of the core competencies. This problem can be easily fixed by increasing the size or using acronyms with a legend. As Navy personnel become more familiar with the competencies, there may be no need to display the legend in the future. Some Vertical group participants also found it hard to find the duty station. In the case of the Horizontal group, participants recommended highlighting changes in rank and reporting senior along with the beginning and end dates for duty stations.

In response to interpreting any piece of information on the display, participants in both groups found it hard to interpret the color changes in the core competency ratings. Similar to the responses to locating information, 53% of the Horizontal group had no difficulty interpreting the information compared to 29% of the Vertical group.

In terms of useful display features, almost all the participants in the Horizontal group pointed to the color coding that highlighted changes in performance across the different metrics. The most common unhelpful design feature for this group was the small

font size and the vast amount of individual performance data shown on a single display. This latter problem may reflect the unfamiliarity of the group with PSRs. Larger font sizes and more color contrast were their main recommendations for change. A few participants suggested removing or adjusting the location of the reporting senior remarks that are at the bottom of the Horizontal display.

Similar to the Horizontal group, the Vertical group also found the color coding along with the core competency and comparative assessment useful features of the display. This group identified few unhelpful features of the display other than the small font size of the core competencies. The most common suggestion was to increase the font size and add more visuals to highlight changes.

While NPS testing subjects were slightly more favorable to the Horizontal display, there were no significant differences in the knowledge check or the promotion scores between the two groups. Apart from increasing the font size and adjusting the color coding to highlight changes in performance over time, this analysis suggests either display would likely meet the requirements of the Navy.

B. MILLINGTON TESTING RESULTS

As described in Section II, we were unable to travel to Millington to run the testing. And the testing was conducted by the Navy Performance and Evaluation team, who recruited the subjects and managed the testing protocol. After receiving hard copies of the knowledge check and surveys, we transcribed the data into an electronic format and conducted the analysis.

Unlike the NPS testing, the Millington testing was split into AM and PM sessions across the two groups, Horizontal and Vertical, based on participant availability.

Unfortunately, this split created unbalance in the number of participants in each group. Table 4 shows that twelve participants were in the AM Horizontal testing group on July 30 and another ten participants were in the PM group on July 31. There were no PM sessions for the Horizontal group on July 31 and vice versa for the AM session on July 31. Only three participants were in the Vertical testing group for the PM session on July 30, and another nine participants were in the vertical group for the AM session on July 31. Thus, the Horizontal group at 24 participants was twice the size of the Vertical at 12 participants.

Apart from the differences in sample size, the two groups were also unbalanced in terms of background and Board experience. As shown in Table 5, almost 60% of the participants in the larger Horizontal group were Navy Officers, another 36% were enlisted Sailors, with civilians accounting for only 5% of the group. In marked contrast, civilians accounted for a quarter of the Vertical group. Enlisted personnel were the largest at 50% of the group and Officers made up another 25%. This imbalance in rank also affected Board experience across the two groups. In the Horizontal group, 70% of participants served on more than two boards and 69% of them had served on four or more boards. In the smaller Vertical group, only five participants (42%) had served on two or more boards. On account of these differences in background and experience, we hesitate to draw strong conclusions on the knowledge check and user experience between the groups as they are likely correlated with rank and experience.

That said, we summarize the results on the knowledge checks and survey in Tables 5 and 6 for completeness. Apart from question 11, which asked participants to identify the number of times Joe Seaman received a 5 across all competencies,

participants in the Vertical group answered more questions correctly than those in the Horizontal group. These differences are not statistically significant, but that may be because of the small samples and imbalance in samples between the groups. Similar to the NPS testing pool, Millington participants answered most of the questions correctly and made errors on the same questions asking about changes in comparative assessment and competency ratings. Based on the responses to these questions at NPS and Millington, we believe a non-trivial number of participants misunderstood these questions. Nonetheless, these responses also suggest that visually highlighting changes in comparative assessment and ratings will enable Board members to quickly process and correctly interpret the data on a PSR.

Despite differences in the number of correct responses to a few questions, participants in both groups assigned similar promotion scores: 85 on average for the Horizontal group and 83 on average for the Vertical group. These averages are remarkably similar to those reported earlier for the NPS participants. An area of difference at Millington was the time to complete the knowledge check. Participants in the Vertical group completed the check a minute and half faster than those in the Horizontal group. This may be related to the extensive board experience of Millington participants, which enabled them to process the information faster, especially on the Vertical display, which is similar to the current PSR used in Selection Boards.

Table 6 summarizes the user experience across the two Millington groups and Figure 6 shows the complete distribution of user responses. While a majority of the participants in both the Horizontal and Vertical groups agreed or strongly agreed with ease of use and relevance of the prototype displays, participants in the Vertical group

were more positive than those in the Horizontal group. Both groups found it easy to locate the necessary information on the displays and found the information itself useful. However, only 77% of participants in the Horizontal group agreed that the color coding of the comparative assessment and competencies was useful compared to 92% in the Vertical group. This translates to five individuals out of twenty-two in the Horizontal group who had neutral or negative views on the color coding compared to one individual out of twelve in the Vertical group. Since both displays used the same color coding, we believe participants conflated the color coding with the orientation of the display. In another difference, 64% of Horizontal group participants agreed that the display helps make more informed decisions compared to 75% of Vertical group participants. As participants in the Horizontal group were more experienced Board Members, we believe the less positive views on the Horizontal display expressed by some of them may reflect their initial discomfort with this new and different way of displaying PSRs. Nonetheless, the same share of participants across both groups were satisfied with the experience of using the prototype displays.

Moving to the free text responses, only 25% of participants in the Vertical group said they experienced difficulty in locating information in the displays compared to almost 50% in the Horizontal group. For the Vertical group, the main issue was locating the reporting senior remarks, while for the Horizontal group, many participants said the initial unfamiliarity with the new visual layout caused confusion. A Horizontal participant suggested reorganizing the most recent evaluation to the left and past evaluations to the right. Participants in both groups identified the comparative

assessment, relative value and competency ratings as useful features in the display. Fewer participants, mostly in the Horizontal group, mentioned reporting senior remarks.

In terms of unhelpful features, some participants identified specific data such as core competency rating while others, especially in the Horizontal group, did not like the color coding, the font size, and the brevity of the reporting senior comment box. One of the Horizontal group participants suggested including additional qualification designation (AQD) and awards. Finally, the most common feedback to improve the display was better color coding to highlight changes and trends in performance over time. Among the Horizontal group, participants also mentioned using a bigger font size, better display of the evaluation dates, and specific comments unique to each participant. Taken together, these results suggest the participants in the Horizontal group had stronger views about changes to the display because it was different from their prior experience with PSRs.

C. DISCUSSION

Due to the many differences in protocol and testing pools between NPS and Millington, we hesitate to draw conclusions on the displays across the two locations. In lieu of a formal comparison, we offer a few thoughts on common themes and differences.

As shown in Tables 2 and 4, participants in both groups and across locations incorrectly responded to questions regarding changes in ratings and comparative assessment over time. Based on the responses, we believe this result is likely related to confusion about those questions rather than systematic issues with interpreting temporal patterns on the displays. Interestingly, the Millington participants outperformed the NPS students on one of the trend questions (Q13), which is likely due to their Selection Board experience as trends in ratings are often part of the deliberations.

Both groups identified color coding as both a helpful and unhelpful design feature. Our interpretation of these responses is that participants across locations and displays want visuals on the displays to reduce their cognitive burden, but they have strong views for and against the specific choice of color palette on the prototype displays. This should definitely be an area of further testing.

With limited Board experience, the NPS testing pool was more favorable to the Horizontal display in contrast to the Millington pool, who favored the Vertical display likely due to their familiarity with the vertical orientation of records in current Selection Boards. Given that the Horizontal and Vertical groups in both locations scored records in the same manner, the Navy can deploy either display without introducing any systematic bias in grading of the record. That said, experienced Officers were less positive about the Horizontal display, and, combined with the other changes to performance metrics, the Navy may want to opt for the Vertical display to minimize the number of changes senior Officers will have to navigate as the new system is rolled out.

IV. CONCLUSION AND RECOMMENDATIONS

Following the recommendations in Bacolod and Helzer (2025) and the Cognitive Performance Group (Celestin et al. 2024), we worked with the Navy Performance and Evaluation team to develop and test two prototype displays for PSRs that may be used in future Navy Selection Boards. We then tested whether the type of displays (Horizontal or Vertical-Traditional) affects the interpretation of performance data and the scoring of a promotion candidate. In a pool of NPS students, we found few significant differences in knowledge checks between the Vertical and Horizontal displays. Both groups assigned near-identical promotion scores to the fictional record of Joe Seaman. However, the NPS participants in the Horizontal group expressed more positive views about their display compared to the Vertical group.

In a pool of Millington participants with more extensive Board experience, we again found few significant differences in knowledge checks across the displays. In that case, participants in the Vertical group expressed more positive statements about their display compared to the Horizontal group. However, this finding is conflated with the extensive Board experience of Vertical group participants and their familiarity with the traditional PSRs in Selection Boards, which are oriented vertically. Yet the two groups again assigned near-identical promotion scores to Joe Seaman, similar to the NPS testing pool.

Based on these findings, we offer the following recommendations as the Navy continues to refine and test changes to the Performance Evaluation System.

1. **Emphasize visual encoding, rather than display orientation, to reduce cognitive burden.**

Across both testing populations, promotion scores assigned to the fictional candidate were nearly identical regardless of display orientation, suggesting that display orientation itself does not meaningfully affect interpretation of performance information. Accordingly, future PSR redesign efforts should focus less on orientation per se and more on visual cues to clearly distinguish high, average, and low performance. These visual encoding recommendations are not specific to the orientation of the display and can be incorporated into the Horizontal or Vertical-Traditional version.

a. **Color coding**: Participants across locations and testing groups identified color coding as a useful visual feature that highlights changes in performance over time and reduces the cognitive burden involved in processing all the information on a PSR. That said, some participants in each location did not like the choice of color palette we used to distinguish between the high and low scores on the displays. This was especially true of the Millington participants, as the color coding was somewhat altered to meet Millington room requirements. We strongly recommend the Navy further test different color palettes in different testing populations before committing to a decision.

b. **Integrate direction visuals**: We also recommend the Navy test different visuals such as arrows, for example, to automatically highlight a sharp increase or decrease in a core competency rating.

c. **Improve legibility**: We recommend using a larger font size for the competencies even if it requires using acronyms and an accompanying legend. Small font size was a common complaint among NPS participants. Improving legibility would reduce visual strain and support efficient interpretation of performance indicators.

d. **Easily locate most recent record**: To take advantage of natural scanning behavior, we recommend the Navy reorient the Horizontal record beginning with the most recent evaluation on the left, also as suggested by multiple Millington participants.

2. **Account for user familiarity and experience in PSR redesign.**

We found that while preferences for display format appear to be shaped by prior exposure and Board experience, there were few differences in knowledge checks, promotion scores, and overall user experience. In the end, the Navy may simply want to move forward with a revamped Vertical display. This may ease the transition to the new performance metrics on the FITREPs/EVALs, particularly for experienced Board members accustomed to legacy formats. If, however, the Navy decides to move forward with the Horizontal display, we recommend they work with seasoned officers to ensure the horizontal PSR captures all the necessary attributes required for a Board deliberation—AQDs, awards, and so on.

3. **Incorporate the new RV metric and comparative assessment.**

By and large, participants in both locations found the new RV metric and comparative assessment useful features. We recommend the Navy continue testing them in the fleet to promote familiarity among Sailors and Officers before the updated FITREPs/EVALs are rolled out. Finally, we recommend training be adopted on the new RV metric, comparative assessment, and on any new promotion board displays.

THIS PAGE INTENTIONALLY LEFT BLANK

FIGURES

FIGURE 1: VERTICAL-TRADITIONAL PROTOTYPE DISPLAY (NPS TESTING)

NAME(LAST, FIRST, MIDDLE): SEAMAN, JOE | Desig/Rate: MM1 | SSN:

PG	STATION	DUTY	DATES	MOS	REPORTING SENIOR NAME	REPORTING SENIOR PG	REPORTING SENIOR TITLE	PROMOTION STATUS	COMPARATIVE ASSESSMENT	REPORT REL VAL	PFA SCORE	CORE COMPETENCY RATING: Learning Mindset & Adaptability	Leadership Skills	Character Development	Judgement & Decision Making	Experience & Competence	Motivation & Drive	REPORTING SENIOR REMARKS
E6	US PACIFIC FLEET	LPO	7-Nov-2023 6 Nov-2022	12	REP SEN6	O6	CO		Exceptionally Ready	96.32	P	4	5	5	5	5	4	Top command performer that received Navy Commendation Medal. A natural innovator, leader, proven motivator and mentor!
		ASST WCS / ACT LPO	7-Nov-2022 6-Nov-2021	12	REP SEN5	O6	CO		Highly Ready	89.92	P	4	5	4	5	4	4	Exceptional First Class Petty Officer performing at the level of Chief Petty Officer. Consummate professional, leader, manager and multi-talented.
		ASST WCS	7-Nov-2021 6-Nov-2020	12	REP SEN5	O6	CO	Recently Promoted			P	4	5	4	4	3	5	A clear choice for advancement to CPO now! Well rounded and thoroughly professional.
E5	US ATLANTIC FLEET	MACHINIST'S MATE / ASST WCS	5-Nov-2020 6-Nov-2019	12	REP SEN4	O5	XO		Highly Ready	83.68	P	4	5	3	4	4	5	Professional knowledge is unparalleled with strict commitment to process improvement. Earned Navy and Marine Corp Medal for outstanding contributions!
			5-Nov-2019 6-Nov-2018	12	REP SEN4	O5	XO		Many Ready	70.74	P	3	4	4	3	3	5	Energetic self-starter, exceptional Petty Officer performing at the level of a seasoned First Class!
		MACHINIST'S MATE	7-Nov-2018 6-Nov-2017	12	REP SEN3	O4	DEPT HEAD		Many Ready	94	P	5	3	3	4	3	4	Professional knowledge is unparalleled. Efforts were vital to successful inspection. Strong recommendation for advancement.
		MACHINIST'S MATE	7-Nov-2017 6-Nov-2016	12	REP SEN2	O4	DEPT HEAD		Many Ready	89.26	P	5	4	4	4	3	4	Committed to the concept of "Service to the Crew" and lives Navy Core Values of Honor, Courage and Commitment.
		MACHINIST'S MATE	7-Nov-2016 6-Nov-2015	12	REP SEN2	O4	DEPT HEAD	Recently Promoted			P	3	4	3	3	3	4	Role model for all Atlantic Fleet Sailors. Needs no supervision. Highly recommended for increased responsibilities!
E4	US ATLANTIC FLEET	MACHINIST'S MATE	5-Nov-2015 6-Nov-2014	12	REP SEN1	E7	DIV CHIEF		Many Ready	83.68	P	4	5	4	4	5	4	Outstanding leadership traits, dynamic and energetic, effectively supervised predeployment onload of stores.
		MACHINIST'S MATE	5-Nov-2014 6-Nov-2013	12	REP SEN1	E7	DIV CHIEF		Many Ready	64.39	P	3	4	3	3	4	2	Self-starter and exceptional role model. Made outstanding contribution to the team in preparation of mid-deployment inspection.

FIGURE 2: HORIZONTAL PROTOTYPE DISPLAY (NPS TESTING)

Last Name, First Name, MI: Seaman, Joe
DODID: 134265978
Current Pay Grade: E-6

Demographics

2013-14	2014-15	2015-16	2016-17	2017-18	2018-19	2019-20	2020-21	2021-22	2022-23
REP SEN1 E7 DIV CHIEF	REP SEN1 E7 DIV CHIEF	REP SEN2 O4 DEPT HEAD	REP SEN2 O4 DEPT HEAD	REP SEN3 O4 DEPT HEAD	REP SEN4 O5 EXEC OFF (XO)	REP SEN4 O5 EXEC OFF (XO)	REP SEN5 O6 COMM OFF (CO)	REP SEN6 O6 COMM OFF (CO)	REP SEN6 O6 COMM OFF (CO)
				Engineering B.S.					
2014	2015	2016	2017	2018	2019.	2020	2021	2022	2023
E-4 Machinists' Mate	E-4 Machinists' Mate	E-5 Machinists' Mate	E-5 Machinists' Mate	E-5 Machinists' Mate	E-5 Machinists' Mate / Assistant WCS	E-5 Machinists' Mate / Assistant WCS	E-6 Assistant WCS	E-6 Assistant WCS/Acting LPO	E-6 LPO
11/6	11/6	11/7	11/7	11/7	11/6	11/6	11/7	11/7	11/7
USS George Washington CVN-73 US Atlantic Fleet			USS George Washington CVN-73 US Atlantic Fleet				USS Comstock LDS-45 US Pacific Fleet		

	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19	2019-20	2020-21	2021-22	2022-23
PROMOTION STATUS			RECENTLY PROMOTED					RECENTLY PROMOTED		
COMPARATIVE ASSESSMENT	MANY READY	MANY READY		MANY READY	MANY READY	MANY READY	HIGHLY READY		HIGHLY READY	EXCEPTIONALLY READY
RELATIVE VALUE	64.39	83.68		89.26	94	70.74	83.68		89.92	96.32
PFA SCORE	P	P	P	P	P	P	P	P	P	P
CORE COMPETENCY										
Learning Mindset and Adaptability	3	4	3	5	5	3	4	4	4	4
Leadership Skills	4	5	4	4	3	4	5	5	5	5
Character Development	3	4	3	4	3	4	3	4	4	5
Judgement and Decision Making	3	4	3	4	4	3	4	4	5	5
Experience and Competence	4	5	3	3	3	3	4	3	4	5
Motivation and Drive	2	4	4	4	4	5	5	5	4	4
REPORTING SENIOR REMARKS	Self-starter and exceptional role model. Made outstanding contribution to the team in preparation of mid-deployment inspection.	Outstanding leadership traits, dynamic and energetic, effectively supervised predeployment onload of stores.	Role model for all Atlantic Fleet Sailors. Needs no supervision. Highly recommended for increased responsibilities!	Committed to the concept of "Service to the Crew" and lives Navy Core Values of Honor, Courage and Commitment.	Professional knowledge is unparalleled. Efforts were vital to successful inspection. Strong recommendation for advancement.	Energetic self-starter, exceptional Petty Officer performing at the level of a seasoned First Class!	Professional knowledge is unparalleled with strict commitment to process improvement. Earned Navy and Marine Corp Medal for outstanding	A clear choice for advancement to CPO now! Well rounded and thoroughly professional.	Exceptional First Class Petty Officer performing at the level of Chief Petty Officer. Consummate professional, leader, manager and multi-talented.	Top command performer that received Navy Commendation Medal. A natural innovator, leader, proven motivator and mentor!

FIGURE 3: VERTICAL-TRADITIONAL PROTOTYPE DISPLAY (MILLINGTON)

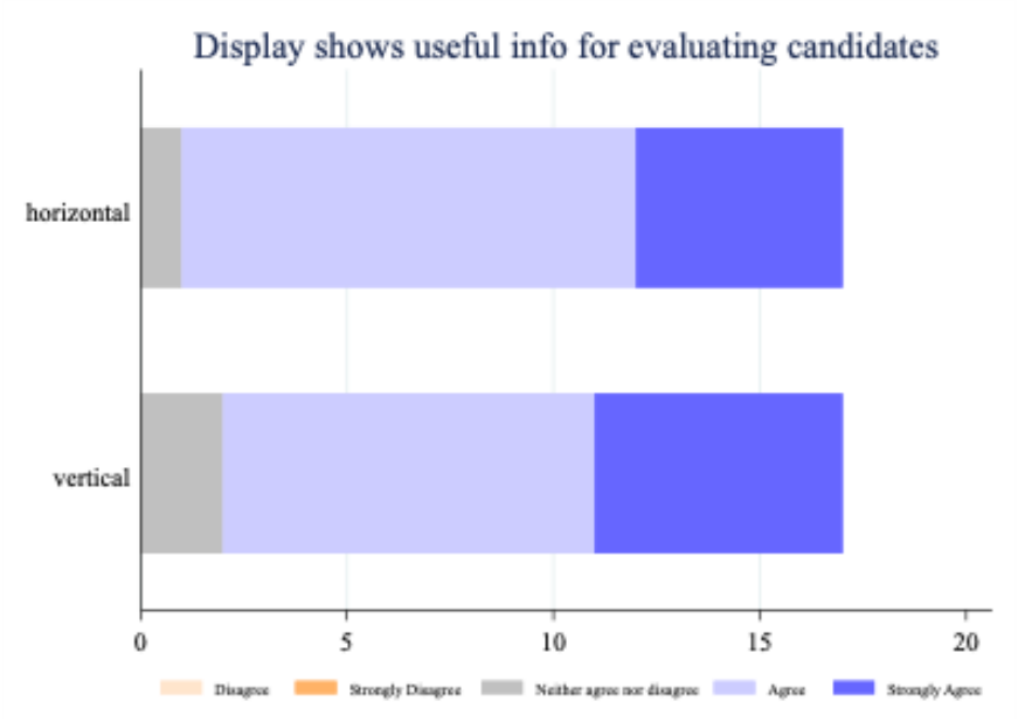
NAME(LAST, FIRST, MIDDLE):			Rate:		DODID:													
SEAMAN, JOE			MM1		8675300001													
PG	STATION	DUTY	DATES	MOS	REPORTING SENIOR NAME	PG	TITLE	COMPARATIVE ASSESSMENT	RELATIVE VALUE	PFA	Learning Mindset & Adaptability	Leadership Skills	Character Development	Judgement & Decision Making	Experience & Competence	Motivation & Drive	REPORTING SENIOR REMARKS
E6	US PACIFIC FLEET	LPO	7-Nov-2023 6-Nov-	12	TARKIN, W. H.	O6	CO	Exceptionally Ready	96.32	P	4	5	5	5	5	4	Top command performer that received Navy Commendation Medal. A natural innovator, leader, proven motivator and mentor!
		ASST WCS / ACT LPO	7-Nov-2022 6-Nov-	12	OBSIDIAN, C. G.	O6	CO	Highly Ready	89.92	P	4	5	4	5	4	4	Exceptional First Class Petty Officer performing at the level of Chief Petty Officer. Consummate professional, leader, manager and multi-talented.
		ASST WCS	7-Nov-2021 6-Nov-	12	OBSIDIAN, C. G.	O6	CO	Recently Promoted		P	4	5	4	4	3	5	A clear choice for advancement to CPO now! Well rounded and thoroughly professional.
E5	US ATLANTIC FLEET	MACHINIST'S MATE / ASST WCS	5-Nov-2020 6-Nov-2018	12	MOGH, W. K.	O5	XO	Highly Ready	83.68	P	4	5	3	4	4	5	Professional knowledge is unparalleled with strict commitment to process improvement. Earned Navy and Marine Corp Achievement Medal for outstanding contributions!
			5-Nov-2019 6-Nov-	12	MOGH, W. K.	O5	XO	Many Ready	70.74	P	3	4	4	3	3	5	Energetic self-starter, exceptional Petty Officer performing at the level of a seasoned First Class!
		MACHINIST'S MATE	7-Nov-2018 6-Nov-	12	BARNES, J. B.	O4	DEPT HEAD	Many Ready	94	P	5	3	3	4	3	4	Professional knowledge is unparalleled. Efforts were vital to successful inspection. Strong recommendation for advancement.
		MACHINIST'S MATE	7-Nov-2017 6-Nov-	12	SMITH, S. A.	O4	DEPT HEAD	Many Ready	89.26	P	5	4	4	4	3	4	Committed to the concept of "Service to the Crew" and lives Navy Core Values of Honor, Courage and Commitment.
		MACHINIST'S MATE	7-Nov-2016 6-Nov-	12	SMITH, S. A.	O4	DEPT HEAD	Recently Promoted		P	3	4	3	3	3	4	Role model for all Atlantic Fleet Sailors. Needs no supervision. Highly recommended for increased responsibilities!
E4	US ATLANTIC FLEET	MACHINIST'S MATE	5-Nov-2015 6-Nov-	12	POPEYE, T. S.	E7	DIV CHIEF	Many Ready	83.68	P	4	5	4	4	5	4	Outstanding leadership traits, dynamic and energetic, effectively supervised predeployment onload of stores.
		MACHINIST'S MATE	5-Nov-2014 6-Nov-	12	POPEYE, T. S.	E7	DIV CHIEF	Many Ready	64.39	P	3	4	3	3	4	2	Self-starter and exceptional role model. Made outstanding contribution to the team in preparation of mid-deployment inspection.

FIGURE 4: HORIZONTAL PROTOTYPE DISPLAY (MILLINGTON)

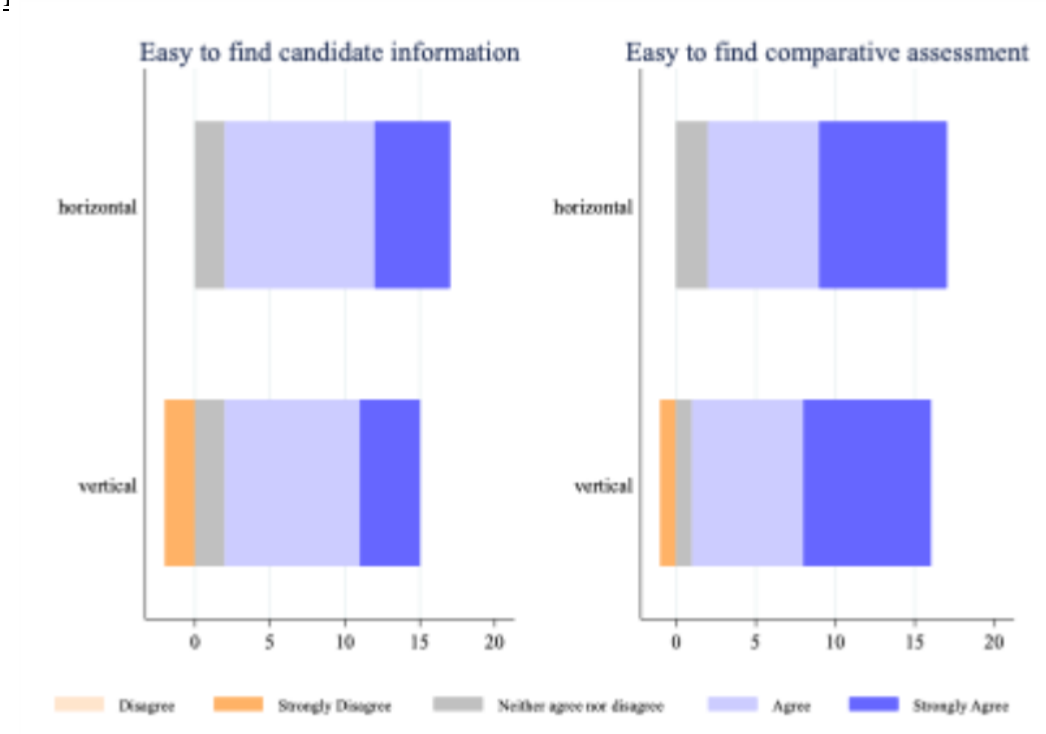
Last Name, First Name, MI: Seaman, Joe			DODID: 134265987					CURRENT PAY GRADE: E-6			
Period Begin	20131116	20141116	20151116	20161116	20171116	20181116	20191116	20201116	20211116	20221116	
Period End	20141115	20151115	20161115	20171115	20181115	20191115	20201115	20211115	20221115	20231115	
EDUCATION:	E-4	E-4	E-5	E-5	E-5	E-5	E-5	E-6	E-6	E-6	
ENGINEERING B.S.	Machinists' Mate	Machinists' Mate	Machinists' Mate	Machinists' Mate	Machinists' Mate	Machinists' Mate / Assistant WCS	Machinists' Mate/ Assistant WCS	Assistant WCS	Assistant WCS/Acting LPO	LPO	
	USS George Washington CVN-73 US Atlantic Fleet		USS George Washington CVN-73 US Atlantic Fleet					USS Comstock LDS-45 US Pacific Fleet			
	REP SEN1 E7 DIV CHIEF		REP SEN2 O4 DEPT HEAD		REP SEN3 O4 DEPT HEAD		REP SEN4 CDR XO	REP SENS CAPT CO			
COMPARATIVE ASSESSMENT	MANY READY	MANY READY	RECENTLY	MANY READY	MANY READY	MANY READY	HIGHLY READY	RECENTLY	HIGHLY READY	EXCEPTIONALLY READY	
RELATIVE VALUE	64.39	83.68	PROMOTED	89.26	94	70.74	83.68	PROMOTED	89.92	96.32	
PFA SCORE	P	P	P	P	P	P	P	P	P	P	
CORE COMPETENCY											
Learning Mindset and Adaptability	3	4	3	5	5	3	4	4	4	4	
Leadership Skills	4	5	4	4	3	4	5	5	5	5	
Character Development	3	4	3	4	3	4	3	4	4	5	
Judgement and Decision Making	3	4	3	4	4	3	4	4	5	5	
Experience and Competence	4	5	3	3	3	3	4	3	4	5	
Motivation and Drive	2	4	4	4	4	5	5	5	4	4	
REPORTING SENIOR REMARKS	Self-starter and exceptional role model. Made outstanding contribution to the team in preparation of mid-deployment inspection.	Outstanding leadership traits, dynamic and energetic, effectively supervised predeployment onload of stores.	Role model for all Atlantic Fleet Sailors. Needs no supervision. Highly recommended for increased responsibilities!	Committed to the concept of "Service to the Crew" and lives Navy Core Values of Honor, Courage and Commitment.	Professional knowledge is unparalleled. Efforts were vital to successful inspection. Strong recommendation for advancement.	Energetic self-starter, exceptional Petty Officer performing at the level of a seasoned First Class!	Professional knowledge is unparalleled with strict commitment to process improvement. Earned NAM for outstanding contributions!	A clear choice for advancement to CPO now! Well rounded and thoroughly professional.	Exceptional First Class Petty Officer performing at the level of Chief Petty Officer. Consummate professional, leader, manager and multi-talented.	Top command performer that received Navy Commendation Medal. A natural innovator, leader, proven motivator and mentor!	

FIGURE 5: NPS USER SURVEY RESPONSES

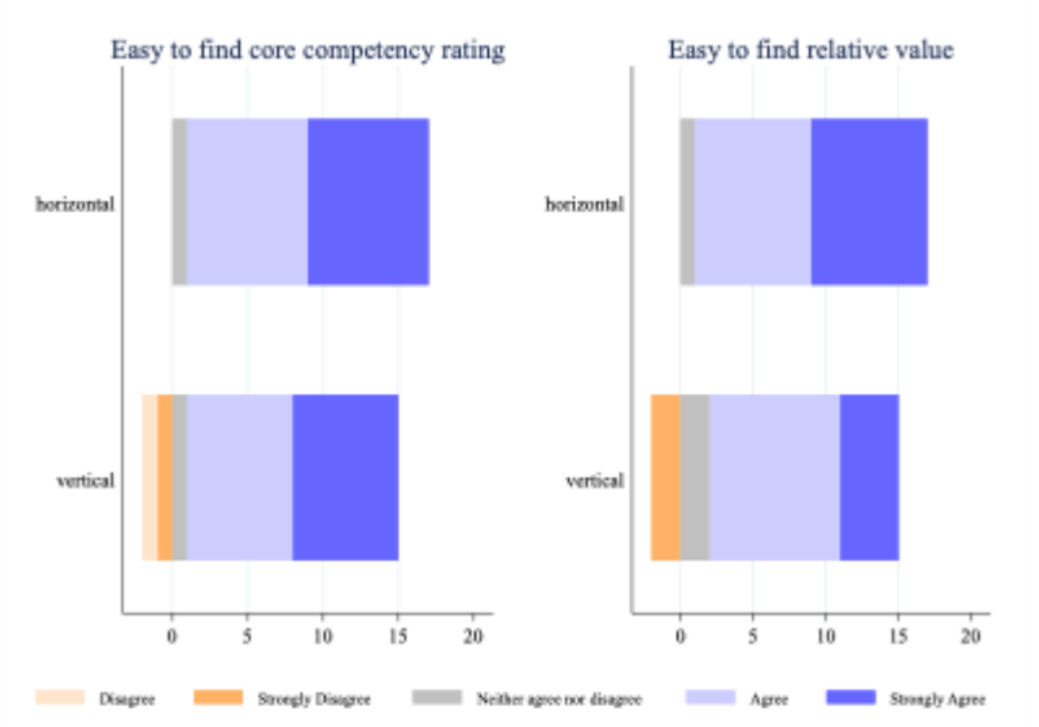
Panel A:



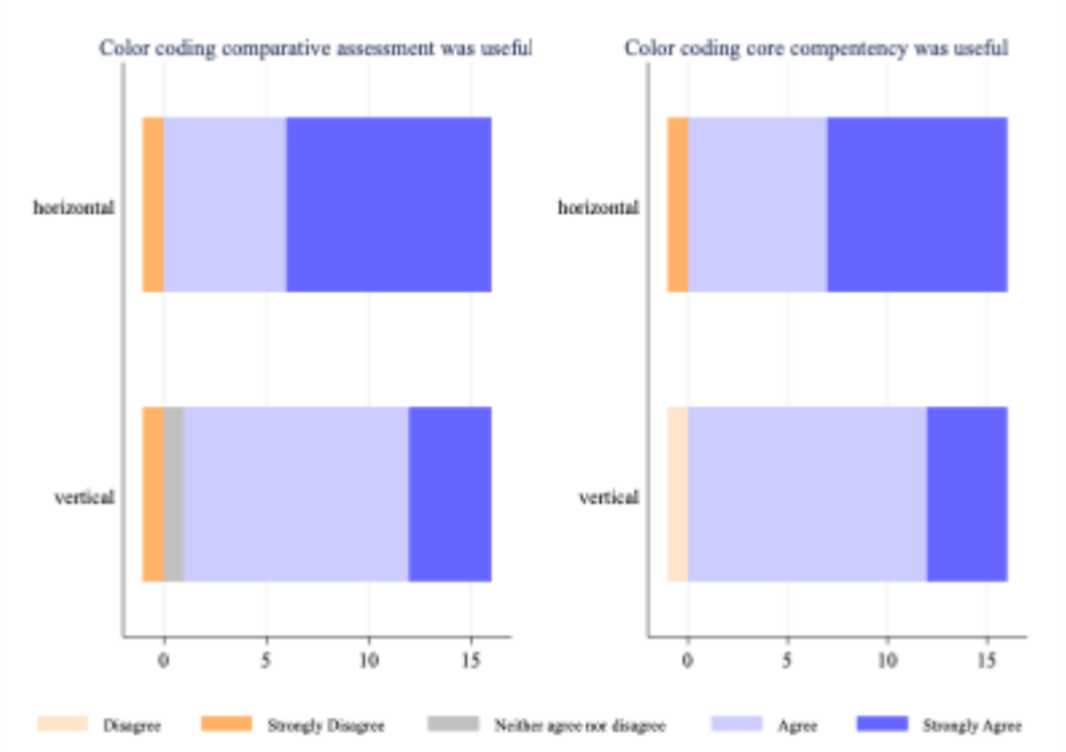
Panel B:



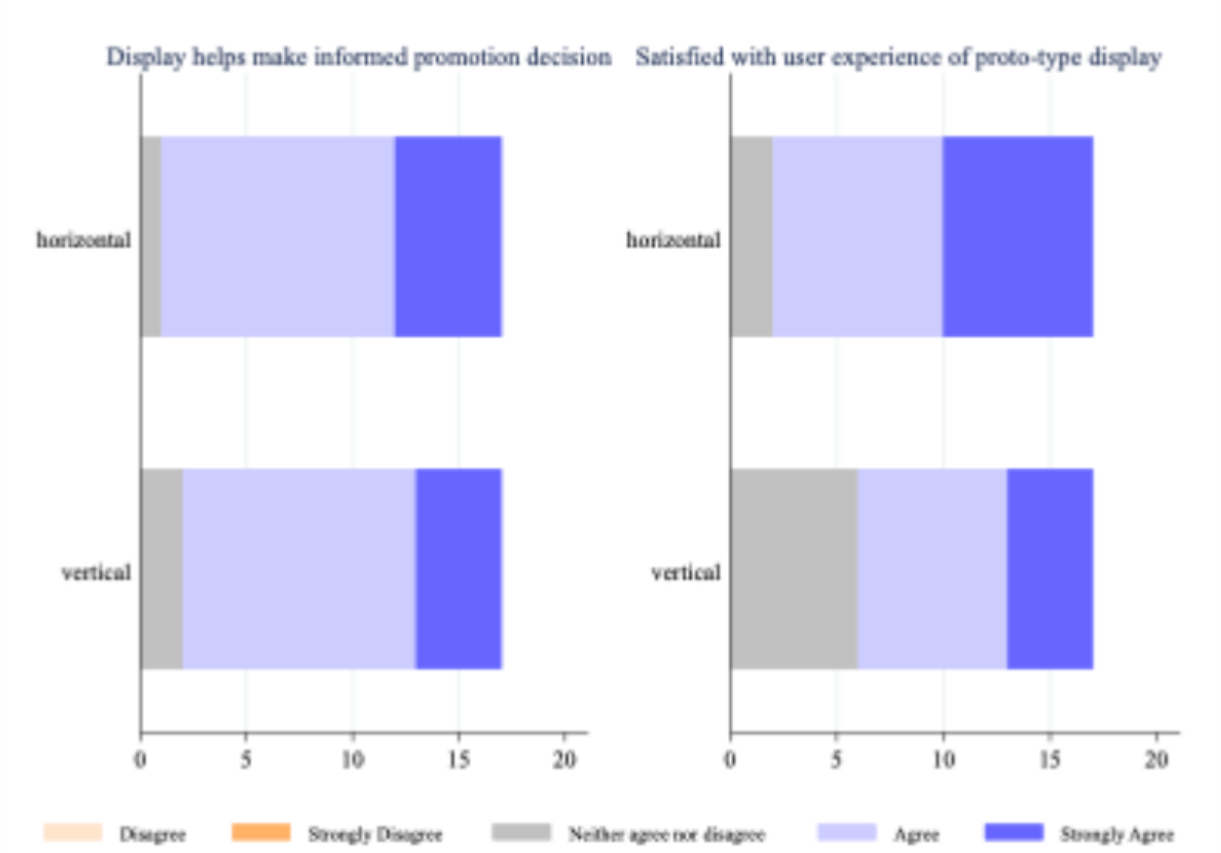
Panel C:



Panel D:



Panel E:



Panel F:

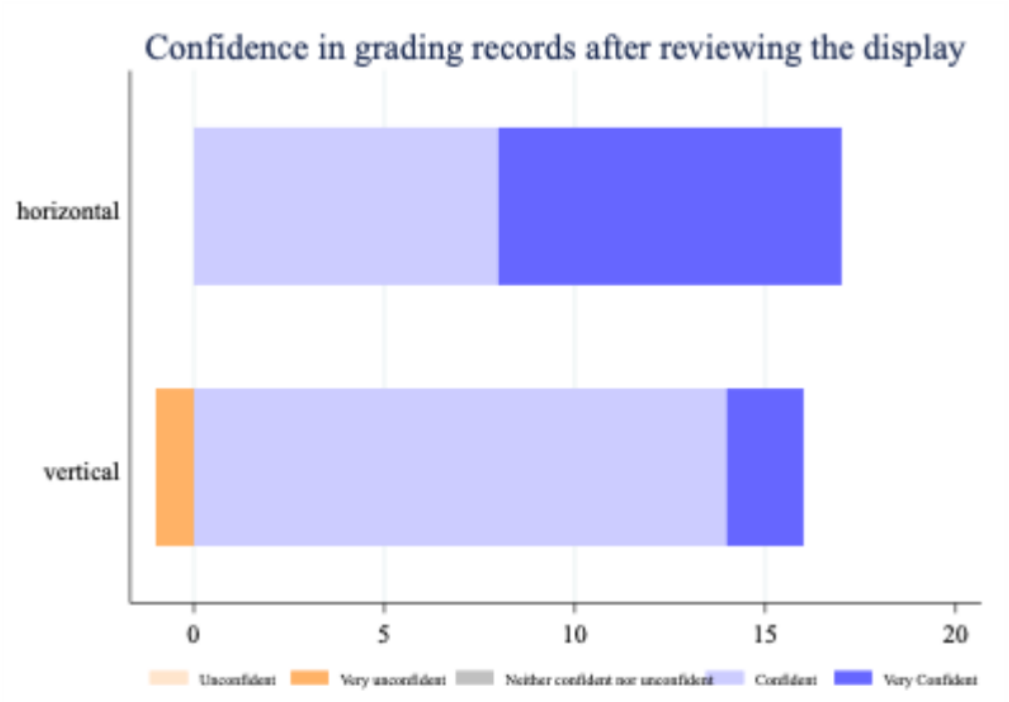
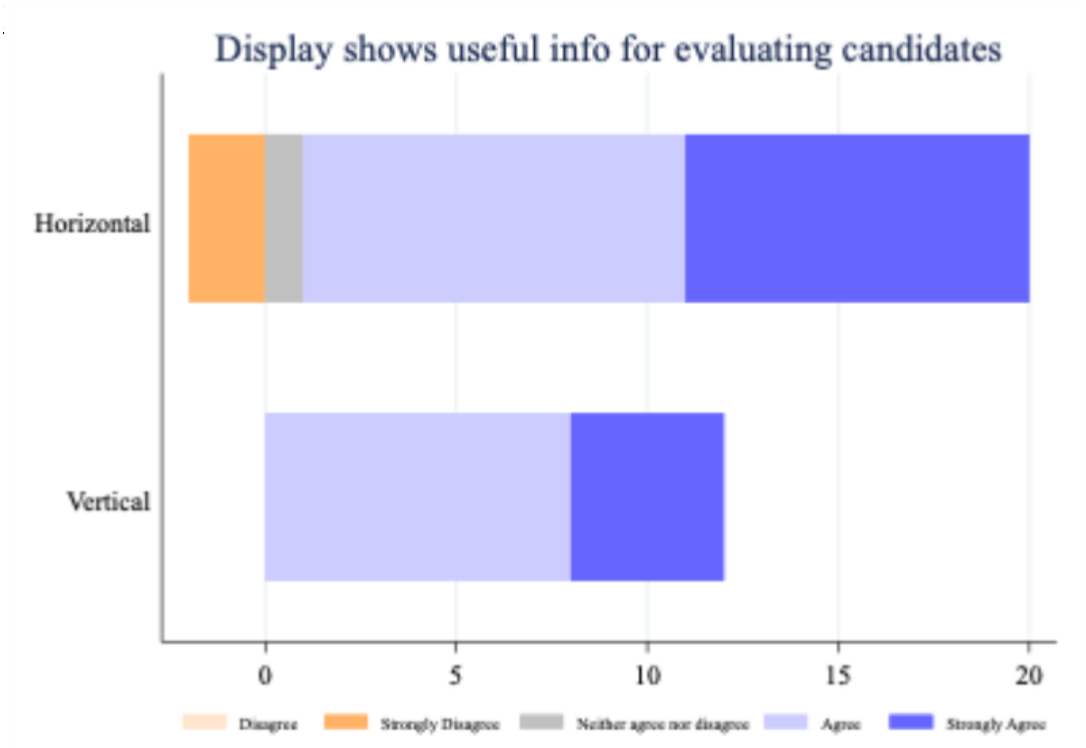
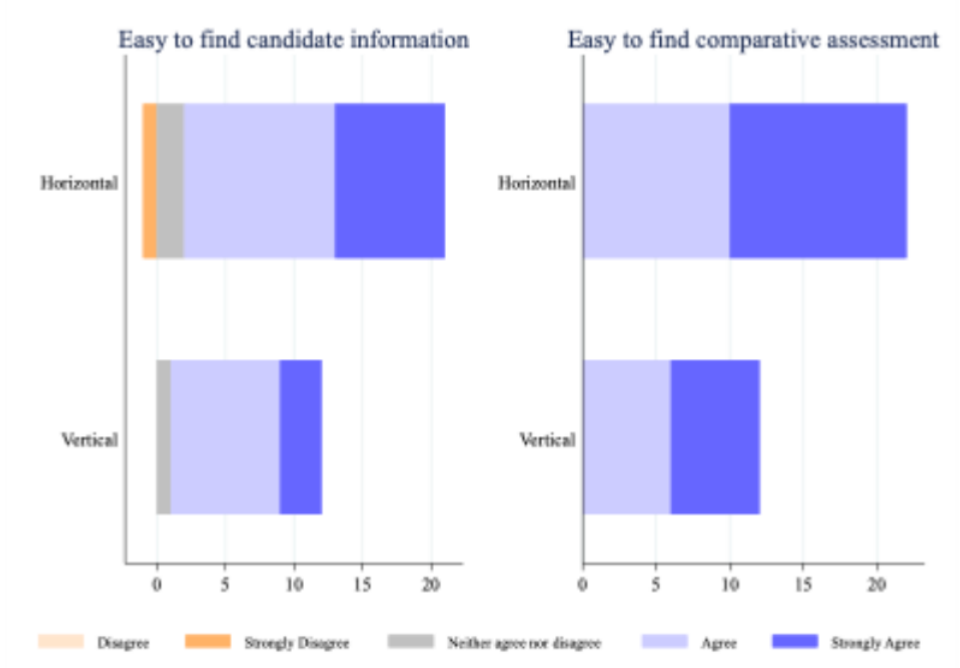


FIGURE 6: MILLINGTON USER SURVEY RESPONSES

Panel A:

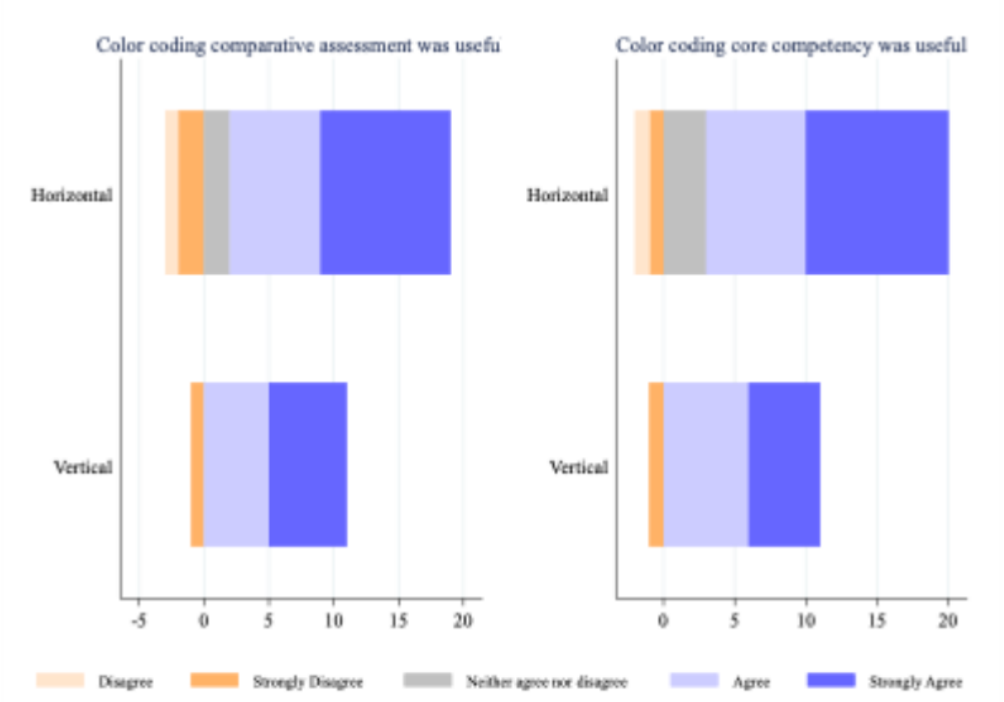


Panel B:

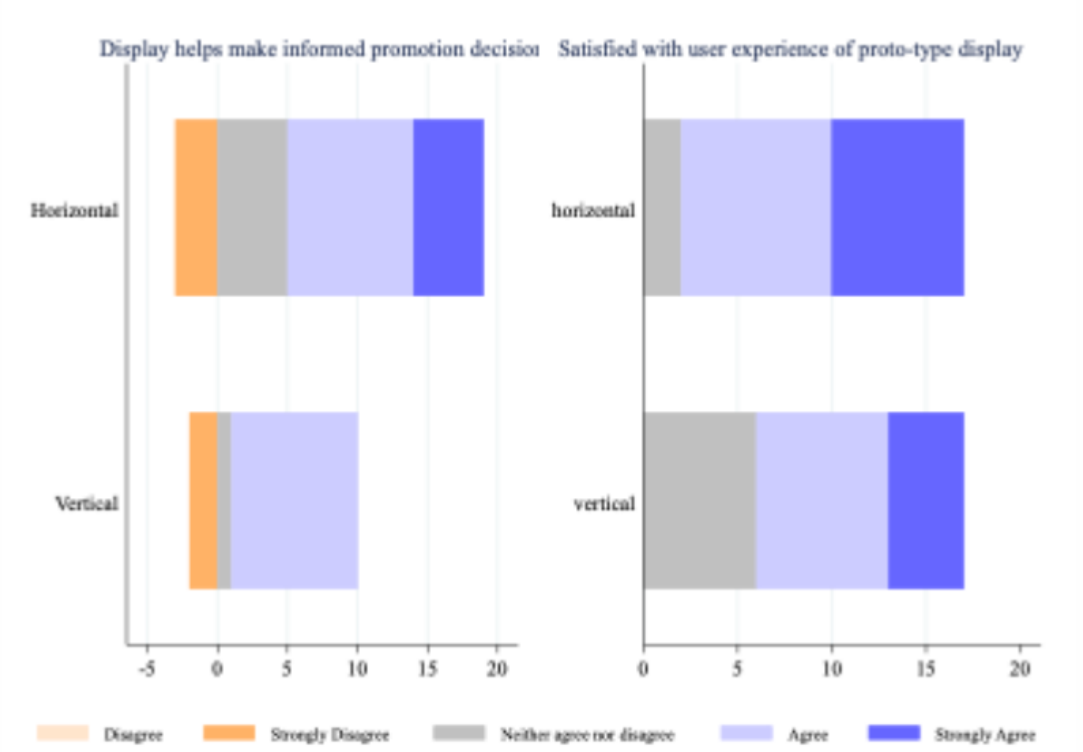


Panel C:

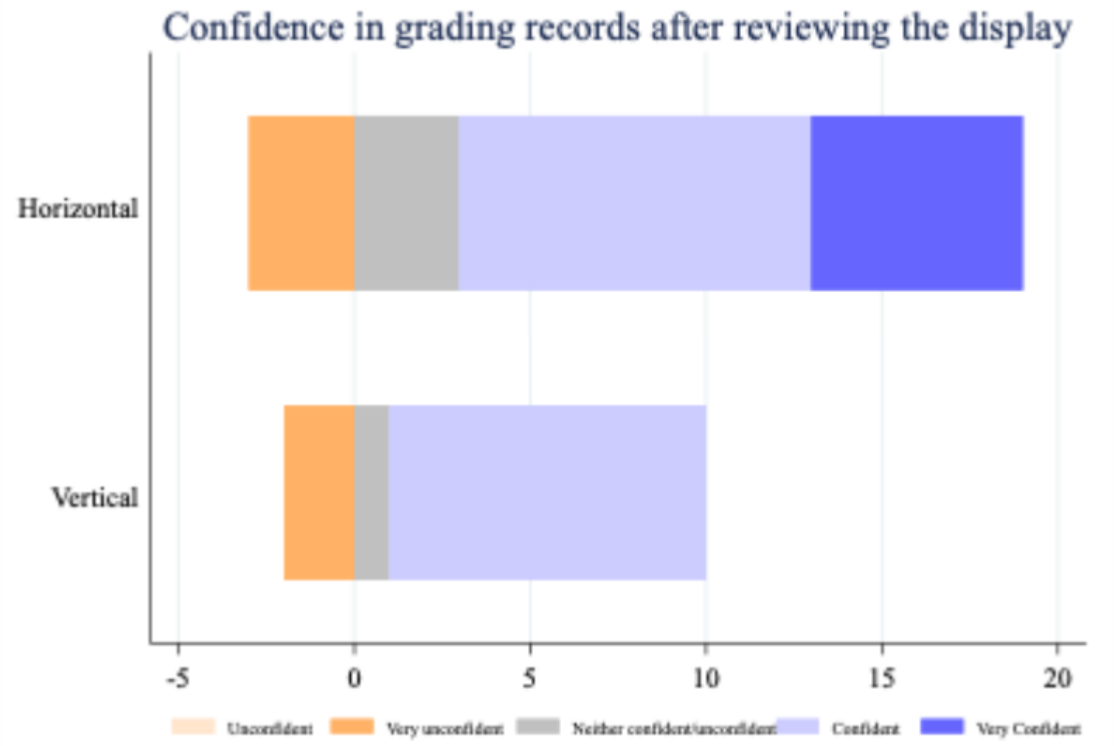
Panel D:



Panel E:



Panel F:



TABLES

TABLE 1: COMPARATIVE ASSESSMENT & CORE COMPETENCIES	
Comparative Assessment	
One of the Few Exceptionally Ready Who Meet the Needs of the Navy	
One of the Highly Ready Who Meet the Needs of the Navy	
One of the Many Ready Who Meet the Needs of the Navy	
Marginally Ready for the Next Pay Grade	
Not Ready	
Core Competency Rating	
1	No potential to perform at next paygrade
2	Some potential to perform at ne›
3	Average potential to perform at next paygrade
4	Above average potential to perform at next paygrade
5	Extremely high potential to perform at next paygrade

TABLE 2: NPS KNOWLEDGE CHECK RESULTS						
	HORIZONTAL		VERTICAL		Diff	t-test (p-value)
	Mean	SD	Mean	SD		
US Navy	41.2%	50.7%	70.6%	49.3%	-29%	0.089*
O3	47.1%	51.4%	47.1%	51.4%	0%	.
O4	35.3%	49.3%	41.2%	50.7%	-6%	0.7338
Selection Board Experience	11.8%	33.2%	11.8%	33.2%	0%	.
ACCURACY & KNOWLEDGE CHECK						
Q1_Duty	1.00	0.00	1.00	0.00	0.00	.
Q2_Station	1.00	0.00	1.00	0.00	0.00	.
Q3_First_Eval_Date	0.88	0.33	0.88	0.33	0.00	.
Q4_RS_Rank	0.88	0.33	0.82	0.39	0.06	0.641
Q5_Relative Value	1.00	0.00	0.94	0.24	0.06	0.332
Q6_Lowest_Comp_Assessment	0.65	0.49	0.76	0.44	-0.12	0.467
Q7_Exceptional_Comp_Assessme	1.00	0.00	1.00	0.00	0.00	.
Q8_Physical_Fitness	1.00	0.00	1.00	0.00	0.00	.
Q9_Core_Comp_Below_2	1.00	0.00	1.00	0.00	0.00	.
Q10_Core_Comp_5 (which CC)	0.94	0.24	0.59	0.51	0.35	0.017**
Q11_Core_Comp_Num_5 (#)	0.53	0.51	0.24	0.44	0.29	0.082*
Q12_Core_Comp_Most_Imp	0.53	0.51	0.71	0.47	-0.18	0.304
Q13_LMA_Trend	0.71	0.47	0.47	0.51	0.24	0.173
Q14_MD_Trend	0.94	0.24	0.76	0.44	0.18	0.158
Q15_Score	85.29	15.46	83.82	15.16	1.47	0.781
Quiz_length_minutes	9.50	1.55	9.65	2.12	-0.15	0.821
*Please see Appendix A for the complete questions on the Accuracy and Knowledge Check						

TABLE 3: NPS USER EXPERIENCE SURVEY RESULTS				
	HORIZONTAL Mean	VERTICAL Mean	Diff	t-test (p-value)
US Navy	41.2%	70.6%	-29%	0.089*
O3	47.1%	47.1%	0%	.
O4	35.3%	41.2%	-6%	0.7338
Selection Board Experience	11.8%	11.8%	0%	.
% Agree or Strongly Agree				
Usefulness and relevance				
Information is useful for evaluating candidate	94.1%	88.2%	5.9%	0.560
Display ease of use				
Easy to find candidate inf	88.2%	76.5%	11.8%	0.384
Easy to find comparative assessment	88.2%	88.2%	0.0%	
Easy to find core compentency ratings	94.1%	82.4%	11.8%	0.303
Easy to find relative value	94.1%	76.5%	17.6%	0.158
Color coding of comp. assessment was useful	94.1%	88.2%	5.9%	0.560
Color coding of core compentencies was useful	94.1%	94.1%	0.0%	
Display helps make more informed board decisions	94.1%	88.2%	5.9%	0.560
Satisfied with the experience of using prototype performance summary display	88.2%	64.7%	23.5%	0.114
% Confident or Very Confident				
In grading of record after display review	100%	94%	6%	0.332

TABLE 4: MILLINGTON TESTING SUBJECTS			
Horizontal			
	AM	PM	Total
30-Jul	12	0	12
31-Jul	0	10	10
Total	12	10	22
Vertical			
	AM	PM	Total
30-Jul	0	3	3
31-Jul	9	0	9
Total	9	3	12

TABLE 5: MILLINGTON KNOWLEDGE CHECK RESULTS						
	HORIZONTAL		VERTICAL		Diff	t-test (p-value)
	Mean	SD	Mean	SD		
Civilian	4.5%	21.3%	25.0%	45.2%	-20.5%	0.162
Enlisted	36.4%	49.2%	50.0%	52.2%	-13.6%	0.466
Officer	59.1%	50.3%	25.0%	45.2%	34.1%	0.055*
Selection Board Experience						
0	9.1%	29.4%	8.3%	28.9%	0.8%	0.943
1	18.2%	39.5%	50.0%	52.2%	-31.8%	0.082*
2 or more	72.7%	45.6%	41.7%	51.5%	31.1%	0.095*
ACCURACY & KNOWLEDGE CHECK						
Q1_Duty	0.91	0.29	0.92	0.29	-0.01	0.943
Q2_Station	0.91	0.29	1.00	0.00	-0.09	0.162
Q3_First_Eval_Date	0.77	0.43	0.92	0.29	-0.14	0.254
Q4_RS_Rank	0.91	0.29	0.92	0.29	-0.01	0.943
Q5_Relative Value	0.73	0.46	0.75	0.45	-0.02	0.890
Q6_Lowest_Comp_Assessment	0.32	0.48	0.58	0.51	-0.27	0.156
Q7_Exceptional_Comp_Assessment	1.00	0.00	0.92	0.29	0.08	0.339
Q8_Physical_Fitness	0.95	0.21	1.00	0.00	-0.05	0.329
Q9_Core_Comp_Below_2	0.95	0.21	0.92	0.29	0.04	0.695
Q10_Core_Comp_5 (which CC)	0.86	0.35	0.92	0.29	-0.05	0.640
Q11_Core_Comp_Num_5 (#)	0.91	0.29	0.75	0.45	0.16	0.288
Q12_Core_Comp_Most_Imp	0.23	0.43	0.50	0.52	-0.27	0.138
Q13_LMA_Trend	0.91	0.29	0.92	0.29	-0.01	0.943
Q14_MD_Trend	0.91	0.29	0.92	0.29	-0.01	0.943
Q15_Score	85.23	14.76	83.33	16.28	1.89	0.741
Quiz_length_minutes	9.00	2.25	7.42	2.50	1.58	0.082*
Number of Participants	22		12			
*Please see Appendix A for the complete questions on the Accuracy and Knowledge Check						

TABLE 6: MILLINGTON USER EXPERIENCE SURVEY RESULTS				
	HORIZONTAL Mean	VERTICAL Mean	Diff	t-test (p-value)
Civilian	4.5%	25.0%	-20.5%	0.162
Enlisted	36.4%	50.0%	-13.6%	0.466
Officer	59.1%	25.0%	34.1%	0.055*
Selection Board Experience				
0	9.1%	8.3%	0.8%	0.943
1	18.2%	50.0%	-31.8%	0.082*
2 or more	72.7%	41.7%	31.1%	0.095*
% Agree or Strongly Agree				
Usefulness and relevance				
Information is useful for evaluating candidate	86.4%	100.0%	-13.6%	0.083
Display ease of use				
Easy to find candidate inf	86.4%	91.7%	-5.3%	0.640
Easy to find comparative assessment	100.0%	100.0%	0.0%	
Easy to find core compentency ratings	90.9%	91.7%	-0.8%	0.943
Easy to find relative value	100.0%	100.0%	0.0%	
Color coding of comp. assessment was useful	77.3%	91.7%	-14.4%	0.254
Color coding of core compentencies was useful	77.3%	91.7%	-14.4%	0.254
Display helps make more informed board decisions	63.6%	75.0%	-11.4%	0.504
Satisfied with the experience of using prototype performance summary display	77.3%	75.0%	2.3%	0.888
% Confident or Very Confident				
In grading of record after display review	73%	75%	-2%	0.890

APPENDIX A: ACCURACY AND KNOWLEDGE CHECK

Please record the current time (quiz beginning time):

Demographics

1. What is Joe Seaman's most recent *duty*?
2. What is Joe Seaman's most recent duty station?
3. In which FY did Joe Seaman receive his first EVAL on this proto-type PSR?

Reporting Senior

4. What is the lowest rank of all reporting seniors to complete an EVAL for JOE SEAMAN?

RS Relative Value

Reporting Senior Relative Value is a numeric comparative assessment of Joe Seaman's performance compared to all other Sailors within his paygrade. Higher numbers represent stronger comparative performance, and lower numbers represent weaker comparative performance.

5. In which FY did Joe Seaman receive his highest relative value?

Comparative Assessment

6. What is the lowest Comparative Assessment Joe Seaman has ever received in this proto-type PSR?
7. In which years has Joe Seaman received an Exceptionally Ready Comparative Assessment?
8. How often has Joe Seaman received a P on his PFA score?

Core Competency Ratings

9. For which core competencies has Joe Seaman received a 2 or below in this proto-type PSR?
10. For which core competencies has Joe Seaman received a 5 in this proto-type PSR?
11. Since 06 Nov 2020, how many times has Joe Seaman received a 5 in any competency?
12. In which core competency has Joe Seaman made the most improvement from 2017-18 to 2022-23?
13. What is Joe Seaman's trend on ratings for Learning Mindset and Adaptability from 2018-19 to 2022-23?
14. What is Joe Seaman's trend on ratings for Motivation and Drive from 2018-19 to 2022-23?

Final Questions

If you were to vote this record at the board, what score would you assign?

0 25 50 75 100

Please record the current time (quiz end time): _____

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B: USER EXPERIENCE SURVEY

We now invite you to complete a brief User Experience survey. As a reminder, participation is voluntary, and all responses will remain completely confidential. Thank you again for your time and insights, which will help to improve the environment that supports promotion and Selection Board members in making informed decisions.

DEMOGRAPHICS

1. What is your current rank?
2. How many Selection Boards have you served on?

USEFULNESS AND RELEVANCE

3. Do you agree or disagree that the prototype display includes information that is useful for evaluating a candidate for promotion?

Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree
○	○	○	○	○

EASE OF USE

4. Do you agree or disagree with the following statements regarding the ease of use of the prototype display?

	Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree
It was easy to find the candidate information on the display.	○	○	○	○	○
It was easy to find the comparative	○	○	○	○	○

assessment on the display.					
It was easy to find the core competency ratings on the display.	○	○	○	○	○
It was easy to find the relative value ratings on the display.	○	○	○	○	○
The color-coding of the comparative assessment was useful.	○	○	○	○	○
The color-coding of the core competencies was useful.	○	○	○	○	○

5. Did you experience any difficulty locating any piece of information on the display?
6. Did you experience any difficulty interpreting any piece of information on the display (comparative assessment or core competency ratings)?

DECISION SUPPORT

7. Do you agree or disagree that the display would help me make more informed promotion board decisions?

Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree
○	○	○	○	○

8. How confident or unconfident did you feel in how you graded the record after reviewing the display?

Very Unconfident	Unconfident	Neither confident nor unconfident	Confident	Very Confident
○	○	○	○	○

SATISFACTION

9. Overall, I am satisfied with the experience of using this prototype performance summary display.

Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree
○	○	○	○	○

OPEN-ENDED QUESTIONS

10. What features of the display did you find helpful?
11. What features of the display did you find unhelpful?
12. What changes would you suggest to improve the display?

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX C: REQUEST TO PARTICIPATE EMAIL (NPS)

Dear MN3070 student,

My name is Prof. Marigee Bacolod, and I'm part of a research team based at Naval Postgraduate School, working in collaboration with the Talent Management Center of Excellence (PERS-00K) at Navy Personnel Command to support improvements to the Navy's performance management system.

We would like to invite you to participate in an evaluation of prototype displays that present Sailor information drawn from Official Military Personnel File (OMPF) and Performance Evaluation Summary Records (PSR), as they would appear at Navy Selection and Promotion Boards. This effort is focused on identifying features of the decision and information environment that may enhance board members' ability to scan, absorb, understand, interpret, and decide during the promotion process.

When: Thursday 17 July, 1500-1600

The testing is expected to take approximately 60 minutes to complete. If you are interested in participating, please respond to this email and say "YES, I would like to participate," and we will follow up with a calendar invitation that includes location details.

We truly value your insights and hope you'll consider contributing to this important initiative.

Sincerely,

Dr Marigee Bacolod and Dr Latika Hartmann

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX D: TESTING SCRIPT OF RECORD

Thank you for agreeing to speak with us today. My name is __ and I am part of a team based at the Naval Postgraduate School working with the Talent Management Center of Excellence (PERS-00K) at Navy Personnel Command to support the improvement of the Navy's performance management system and how information is displayed at Selection Boards.

[*For Millington test subjects, modify to "*My name is __ and I am part of the Talent Management Center of Excellence (PERS-00K) at Navy Personnel Command. The Navy Performance Management Administration team has been working for the past several years to develop policy and technology that will support the next generation of performance management in the Navy."]

For all subjects: The future system will provide more accurate information for Selection Boards to differentiate Sailors and improve their ability to identify the best and most fully qualified for both administrative and statutory boards. In order to achieve these goals, we are testing policy updates that will: 1) identify reporting seniors lower in the chain of command who have a clearer understanding of Sailor performance and 2) provide opportunities beyond RSCA for reporting seniors to send clear performance ratings to the board. A critical part of this process is to understand how this new, expanded information will be displayed to Selection Board members in a way that is easy to understand.

To support this effort, we are seeking feedback from you on prototype displays that present hypothetical Sailor information that could be contained in their Official Military Personnel File (OMPF) and Performance Summary Records (PSR) at the Navy Selection and Promotion Boards when this new policy is approved. In particular, we seek to promote features of the decision/information environment that may improve Selection Board members' ability to scan, absorb, understand, interpret, and decide in the Tank voting process.

We estimate it will take no more than 60 minutes to complete the exercise. Thank you for your participation.

ROUTINE USES: Your responses will be combined with the responses of all others and will not be attributed to any single individual. The anonymized responses will be stored on a password-protected server at the Naval Postgraduate School.

CONFIDENTIALITY: All responses will be kept COMPLETELY confidential. All responses will be coded and statistically summarized and will not be attributed to any single individual. Individual responses will not be shared with Navy Personnel Command.

PARTICIPATION: Participation is entirely voluntary. Failure to participate and/or respond to any of the quiz or survey questions will NOT result in any penalties except

possible lack of representation of your views in the final results. You may withdraw your participation at any time.

Thank you for agreeing to be part of our study!

Part I. Background *(5-10 mins)*
[At NPS, Marigee/Latika will be in separate classrooms. One classroom will see the horizontal display; the other classroom the vertical display.]
[For Millington test subjects, each group will see either the horizontal or the vertical display based on the following schedule: Wed AM (horizontal); Wed PM (vertical); Thurs AM (vertical); Thurs PM (horizontal)]

BEFORE TURNING ON THE DISPLAY: Before we show you the prototype display, we will play a short video that will explain how reporting seniors will rate Sailors under the hypothetical future policy. Please note, these changes to policy are still undergoing rigorous testing and have not been fully approved by CNP. You will have an opportunity to provide any feedback you have about the proposed policy changes at the end of the study. Please keep in mind that the purpose of this study is to provide feedback on the displays using the information as it is explained in this video. Does anyone have any questions before I play the video?
[Play the video explaining the data points in SPAR Module 4.]

Part II. Administer the Accuracy and Knowledge Check Quiz [Metrics sought: accuracy, time] *(10-15 mins)*
Does anyone have any questions about anything they saw in the video before we project the display showing a hypothetical Sailor's performance summary report that would be shown at a Navy Promotion and Selection Board?
[Pause here and answer any questions]

[*For Millington test subjects ….*]
In a moment, we will project the prototype display onto the screen. You each should have two packets labeled "Packet 1" and "Packet 2" at your chair and something to write with [*provide a pen if anyone doesn't have one].* Please open to the first page of Packet 1 [*pause]*.
Packet 1 contains several questions for you to answer based on the display you will see on the screen. Please complete the questions as quickly and accurately as possible. When you are finished, please close the packet and raise your hand and a member of our team will collect your responses. You do not need to put your name or any identifying information on your packet, just please note the time you ended on your paper. Does anyone have any questions before we begin?
[*Display proto-type, see schedule for visual. Note the start time! Start timer (this will give us an idea of time bounds, in case they forget to note end time).]*
[Wait for participants to complete Packet 1; when all participants have finished Packet 1, continue with the script; Stop timer.]
[*Wait for participants to complete Packet 2; when all participants are finished, continue with the script.]*

*****

[*For NPS test subjects, instead of Packets, instructions are ….*] We will now project the prototype display onto the screen. If you will please open your laptops and navigate to the MS Forms link we sent to your emails, please complete the questions as quickly and accurately as possible. Please answer the questions using the information on the display. Remember, this is not an evaluation of your own process, but simply for us to learn the most efficient way to display information to the Boards. When you are finished, please navigate to the second MS Forms link for the User Experience Survey.
[*For Millington test subjects, they will fill out and hand-write all responses.]*

Part III. Survey [Metrics sought: user preference, overall perceptions, confidence in decision-making] *(10 mins)*
[*For Millington test subjects]* Now, please open your second packet and complete the short survey in Packet 2. When you are done, please raise your hand and a member of our team will collect your response. Once everyone is done, we will finish with a short focus group, where we will solicit any additional feedback on your experience with the display and information presented.

Part IV. Focus Group

[*For NPS test subjects*] We will now be recording our open-ended discussion. Do you consent to having this session recorded and transcribed for the purpose of data analysis and coding? *(If no one objects, hit a record button on iPhone, or if AV-capable classroom, turn on mics and record via podium)*

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Ahn, S. Y., and L. Hartmann. (2022). *Performance Evaluation Needs Assessment* (Tech Rep. No. NPS-22-N100-A). Naval Postgraduate School.

Ahn, S. Y., and L. Hartmann. (2023). *Determinants of Navy Performance Evaluation Records* (Tech. Rep. No. NPS-23_N102-A). Naval Postgraduate School.

Ahn, S. Y., and L. Hartmann. (2025). *Identifying Cognitive Biases in Selection Boards Using a Data-Driven Approach* (Tech. Rep. No. NPS-24-N063-A). Naval Postgraduate School.

Bacolod, M., and E. Helzer. (2022). *Performance Trait Validation* (Tech. Rep. No. NPS-22-N137-A). Naval Postgraduate School.

Bacolod, M., and E. Helzer. (2023). *Phase II, Performance Trait Validation* (Tech. Rep. No. NPS-23-N194-A). Naval Postgraduate School.

Bacolod, M., and E. Helzer. (2025). *Performance Evaluation and Promotion Recommendation: Best Practices for Selection Boards* (Tech. Rep. No. NPS-24-N064-A). Naval Postgraduate School.

Celestin, B. D., T. N. Daly, H. C. Baxter, J. S. W. Spinda, and D. H. Heath. (2024). *Assessing and Developing Data Visualizations for Selection Board Decision-Making* (Final Technical Report). Cognitive Performance Group.

Corbett, A. A. (2024). *Navy Promotion Decision-making* [Master's thesis, Naval Postgraduate School]. NPS Archive: Calhoun. https://hdl.handle.net/10945/73090

Hill, M. (2025). *The Right Stuff: Enhancing Navy and Marine Corps Promotion Boards* [Master's thesis, Naval Postgraduate School]. NPS Archive: Calhoun. https://calhoun.nps.edu/entities/publication/6045880d-de71-4341-ac0b-6c6010d3c168.

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus & Giroux.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Naval Research Program
   Naval Postgraduate School
   Monterey, CA