Usability Results from Human Performance Feedback and Development (HPFD) and ePerformance System Users: Technical Report #1

FINAL REPORT
Submitted to

Navy Human Performance Center (HPC – N5)
ATTN: CDR Mark J. Bourne, MSC, USN
Dam Neck Annex
2025 Tartar Ave, Suite 100
Virginia Beach, VA, 23461-1924

Submitted by

Elizabeth Dean, M.A.
Kimberly R. Aspinwall, M.A.
Michael J. Schwerin, Ph.D.
Douglas E. Kendrick, M.S.

RTI International*
3040 Cornwallis Road
Research Triangle Park, NC 27709-2194

Telephone: (919) 316-3878
Fax: (919) 541-1261
E-mail: schwerin@rti.org

February 15, 2006

RTI
INTERNATIONAL

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE *(DD-MM-YYYY)* 02-15-2006	2. REPORT TYPE Technical Report	3. DATES COVERED *(From - To)* January 2005 – February 2006

4. TITLE AND SUBTITLE

Usability Results from Human Performance Feedback and Development (HPFD) and ePerformance System Users: Technical Report #1

5a. CONTRACT NUMBER
GS-10F-009L7

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Elizabeth Dean, Kimberly Robbins, Michael Schwerin, Douglas E. Kendrick, Mark Bourne

5d. PROJECT NUMBER
0209482

5e. TASK NUMBER
006

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) AND ADDRESS(ES)

RTI International
3040 Cornwallis Road
PO Box 12194
Research Triangle Park, NC 27709

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Navy Human Performance Center
Dam Neck Annex
2025 Tartar Ave, Suite 100
Virginia Beach, VA, 23461-1924

10. SPONSOR/MONITOR'S ACRONYM(S)
HPC-N5

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
A – Approved for public release; distribution is unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The Navy Human Performance Center (HPC) and Navy Personnel Command (NPC) have reached the final phase of development for the Navy's Human Performance Feedback and Development (HPFD) and ePerformance system. This second and final iteration of usability testing was designed to verify the functionality of system revisions made as a result of recommendations offered in Schwerin et al. (2004) and gather user perceptions of system usability and cultural issues that may affect HPFD and ePerformance system implementation. This report documents the results of usability test and user survey data collection. Usability timing and errors analyses suggest that system enhancements as a result of the first usability study improved the overall usability of the HPFD and ePerformance system, but users still have significant difficulty finding, opening and creating HPFD and ePerformance documents.

15. SUBJECT TERMS
HPFD, ePerformance, Navy, usability testing, Human Systems Integration (HSI)

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Michael Schwerin
a. REPORT UNCLASSIFIED	**b. ABSTRACT** UNCLASSIFIED	**c. THIS PAGE** UNCLASSIFIED	UNLIMITED	56	**19b. TELEPHONE NUMBER** *(include area code)* 919-316-3878

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Table of Contents

List of Tables

Executive Summary

The Navy Human Performance Center (HPC) and Navy Personnel Command (NPC) have reached the final phase of development for the Navy's Human Performance Feedback and Development (HPFD) and ePerformance systems. This second and final iteration of usability testing was designed to verify the functionality of system revisions made as a result of recommendations offered in the first round of usability testing and gather user perceptions of system usability and cultural issues that may affect HPFD and ePerformance system implementation.

Usability testing was conducted at three Navy locations—Naval Meteorology and Oceanographic Center (NAVMETOCCEN) Norfolk in Norfolk, Virginia; USS HOWARD (DDG 83) in San Diego, California; and the Bureau of Naval Personnel (BUPERS) in Arlington, Virginia. Thirty-four officer and enlisted supervisors and nonsupervisors took part in usability testing. The most notable findings from this study centered on an overall reduction in user burden compared to previous usability testing but continued difficulty locating and opening the correct HPFD and ePerformance documents. Key findings and observations from this research are listed below.

- HPFD tasks were completed more quickly and with fewer errors than had been observed in previous research.

- Users of HPFD and ePerformance systems required the Quick Reference Guide (QRG) to locate, open, and create documents.

- Tasks were completed much more efficiently and with fewer errors when using a shipboard NSIPS-dedicated NMCI server.

- Supervisors and nonsupervisors experienced similar user problems.

- Overall, users were more satisfied with the professionalism, efficiency, and overall effectiveness of the revised HPFD and ePerformance systems compared to the previous versions.

These positive findings represent a dramatic departure from previous usability research, reflecting system enhancements made prior to the current usability study and the introduction of

usability testing aboard a ship with an onboard NSIPS server. Study results and conclusions lead to the following recommendations:

- Continue to monitor NSIPS server and NMCI connection speeds to maintain system efficiency. Server and connection efficiency are critical to system functionality and user satisfaction.

- Plan to implement a QRG with detailed instructions on how to locate, open, create, and complete performance management and appraisal documents.

- Implement a full-scale pilot study of the HPFD and ePerformance systems to provide a more generalizable evaluation of the system.

- Continue comprehensive system testing with both nonsupervisors and supervisors.

1 Introduction

As Navy Personnel Command (NPC) develops a new performance management and appraisal system, it is confronted with having to develop a system that is fully operational and integrated with the performance evaluation and promotion selection cycle. The Navy Human Performance Center (HPC) and NPC have reached the final phase of development for the Navy's Human Performance Feedback and Development (HPFD) and ePerformance systems. Previous research has (1) identified supervisory and nonsupervisory performance dimensions (Hedge, Borman, Bruskiewicz, & Bourne, 2002); (2) developed behavioral rating anchors using examples of workplace behaviors that reflect outstanding, average, and substandard workplace performance (Borman, Hedge, Bruskiewicz, & Bourne, 2003); (3) conducted focus group interviews among senior leaders to weight the various performance dimensions (Hedge, Bruskiewicz, Borman, & Bourne, 2004); and (4) conducted preliminary usability testing of the HPFD and ePerformance systems (Dean, Schwerin, Lee, Robbins, & Bourne, 2004; Schwerin, Dean, Robbins, & Bourne, 2004). This report documents the second and final iteration of usability testing, designed to (1) verify the functionality of system revisions made as a result of recommendations offered in Schwerin et al. (2004) and (2) gather user perceptions of system usability and cultural issues that may affect HPFD and ePerformance system implementation. This thorough development and testing process assures that Sailors and Fleet customers will have a performance appraisal process that possesses "best practice" features of industry and military performance appraisal processes and has undergone the thorough Human Systems Integration (HSI) testing required by DoD Instruction 5000.2, recently issued by the Undersecretary of Defense for Acquisition, Technology, and Logistics (USD AT&L) (Department of Defense [DoD], 2003).

Within the Department of Defense (DoD) and the Department of the Navy (DoN), there is a growing emphasis on the importance of human systems integration (HSI) in the development of new systems for military personnel. Although HSI evaluations are routinely integrated into training systems (Buff, 2004), it unclear whether HSI is a critical part of the system development process for manpower and personnel systems. DoD Instruction 5000.2 specifically calls for DoD acquisition program managers to "… ensure human factors engineering/cognitive engineering is employed during systems engineering over the life of the program to provide for effective

human-machine interfaces and to meet HSI requirements. Where practicable and cost effective, system designs shall minimize or eliminate system characteristics that require excessive cognitive, physical, or sensory skills; entail extensive training or workload-intensive tasks; result in mission-critical errors; or produce safety or health hazards" (DoD, 2003, enclosure 7, paragraph E7.1.1, p 43). It is clear that it is the DoD's intent to ensure that all systems with a human-machine interface—including manpower and personnel systems—are tested for ease of use and that acquisition program managers need to consider system usability through the life cycle of system development.

To develop a performance appraisal system that provides more accurate assessments of workplace performance, is easy to use, and retains desired aspects of Navy culture (e.g., face-to-face performance feedback and appraisal, focus on the operational mission, and accountability for all supervisors in the chain-of-command), both qualitative and quantitative data on system usability were collected. HPFD and ePerformance system usability testing yielded quantitative and qualitative data about the user and system errors, reactions to the system, and self-reported satisfaction with the system from users. Focus group interviews yielded additional qualitative data reflecting the attitudes, perceptions, and opinions of HPFD and ePerformance system users on the performance appraisal system, potential barriers to implementation, and recommendations to facilitate Fleet-wide implementation (see the companion report by Schwerin, Dean, and Aspinwall [2006] for details). Navy HPFD and ePerformance system implementation managers can use these data to identify next steps in (1) determining system and resource requirements, (2) planning additional system testing and evaluation (e.g., load testing), (3) developing policy, and (4) defining user training requirements. This technical report summarizes the findings of the usability testing in order to assist system managers in identifying the next steps in performance appraisal system development.

2 Literature Review

2.1 Performance Appraisal Systems

As NPC develops a new performance management and appraisal system it must also consider the impact the system will have on various organizational outcomes. Performance appraisal systems need to capture workplace performance in a way that instills feelings of trust, fairness, and confidence in the organization. The impact of performance appraisal systems on job satisfaction, commitment to the organization, and employee retention has been a topic of research among civilian researchers for several years. This research has consistently demonstrated that performance appraisal fairness is the critical element in building a workplace where employees are satisfied with their jobs and committed to the organization.

A strong statistical relationship has been observed repeatedly between work life factors, including measures of satisfaction with the performance appraisal process, and job satisfaction among those in the civilian workforce (Blau, 1999; Ellickson & Jogsdon, 2002). Diekmann, Barsness, and Sondak (2004) found that the more uncertain an employee is about performance standards, the stronger the relationship between his or her perceptions of fairness and job satisfaction. Mayer and Davis (1999) observed that employees' trust for top management and the organization increased when a performance appraisal system that was perceived as inaccurate and unfair was replaced by a new performance appraisal system. After controlling for actual performance ratings, Levy and Williams (1998), demonstrated that performance appraisal satisfaction and perceived system knowledge have a strong, significant relationship with both job satisfaction and commitment to the organization.

An additional body of research has demonstrated statistical ties between perceptions of fairness and commitment to the organization. Daily and Kirk (1992) examined perceptions of workplace/procedural fairness and demonstrated a strong relationship between workplace fairness measures (such as procedural justice and satisfaction with the performance appraisal process) and voluntary turnover intent. When examining the effect of work factors on retention plans, after controlling for the effects of demographic characteristics and distributive justice, Jones (1998) found that the perceived fairness of procedures for pay determination, performance appraisals, and appeals were related to voluntary turnover. Commitment to the organization is

positively associated with employee perceptions of general fairness (Masterson, 2001) and fairness of the performance appraisal system and promotion process (Lemons and Jones, 2001). Simons and McLean Parks (2000) showed that fair treatment of employees positively affects employee retention and commitment to the organization. Giles, Findley, and Field (1997) found that the greatest impact on employees' perceptions of appraisal fairness came from performance system formality and commitment.

Other studies have explored Sailor satisfaction with the Navy performance appraisal system. Descriptive analyses of data from the 2000 Navy-wide Personnel Survey indicates that most Sailors understand the performance appraisal, advancement, and promotion systems, but fewer believe that the most deserving Sailors receive the highest ratings on annual performance appraisals (Olmsted & Underhill, 2003). Although over half of enlisted personnel (58 percent) and over three fourths of officers (77 percent) believed their current performance appraisal system was "fair and accurate," only 29 percent of enlisted personnel and 49 percent of officers believed that "the most qualified and deserving Sailors rank high on their EVAL/FITREP." [1] In addition, although a majority of enlisted personnel and officers reported that they understand the advancement and promotion system (76 percent of enlisted personnel; 83 percent of officers), only 31 percent of enlisted personnel and 50 percent of officers were "satisfied with the present Navy advancement and promotion system." Only 20 percent of enlisted personnel and 41 percent of officers believed that "the most qualified and deserving Sailors get advanced or promoted." These results seem to suggest a low to moderate satisfaction with the Navy's current performance management and appraisal systems.

In 2001, the Navy began development of a new behaviorally based performance appraisal and management system (Executive Review of Navy Training, 2001). The first study in this system development program (Hedge et al., 2002) resulted in the development of the HPFD model—a behaviorally based job performance management system with dimensions for supervisory and nonsupervisory personnel that reflect the qualities that Navy leaders endorse as essential for maintaining a high-quality Navy workforce. Subsequent research (Borman et al., 2003; Hedge et al., 2004) identified the relative strength of performance dimensions at career stages for Navy enlisted personnel (recruit-apprentice, apprentice-journeyman, and journeyman-

---

[1] "EVAL" refers to performance evaluations generated for Sailors in the E1–E6 paygrades. "FITREP" refers to fitness reports generated for Navy personnel in the E7–E9 and O1–O9 paygrades.

master) and officers (junior, mid-grade, and senior). This work culminated in the development of a Web-based HPFD and ePerformance system situated within the Navy Standard Integrated Personnel Systems (NSIPS) for test and evaluation.

2.2 Usability Testing

This study examines the usability of the Navy's pilot HPFD and ePerformance appraisal systems. In theory, new automated systems should reduce the burden on users; in practice, however, such tools can be more difficult to use than their paper counterparts. HSI studies and usability testing (Dumas & Redish, 1993) assess the ease of use, sources of system errors, cultural barriers, and other process concerns that could impede system implementation, acceptance, and use. The research literature calls for usability testing to be conducted using an iterative approach, preferably on site, in conditions that are similar to those the user would actually encounter when interacting with a system. Nielsen (1993, p. 165) describes usability testing as the most fundamental method of system testing and "irreplaceable" because it is the only mechanism that allows the researcher to obtain direct, detailed information on the user's experience with the product being tested. Identifying these potential sources of burden and reducing the causes of user stress can result in a more efficient Web-based performance management and appraisal system.

2.3 Summary of Results from Usability Study I

Results from usability testing and focus group interviews concerning the HPFD and ePerformance systems reported in Schwerin et al. (2004) and Dean et al. (2004) indicate that users noted concerns in two areas—system functionality and performance appraisal process/military culture concerns. Both areas of concern negatively affected user perceptions of the system, and recommendations were made to facilitate system improvement.

The primary concern for users was system performance and the functionality of proposed features of the HPFD and ePerformance systems. Specifically, the frequency and severity of connectivity problems with the Navy Marine Corps Intranet (NMCI), which is used to access the NSIPS server, had a significant negative effect on user perceptions of the two systems. This negative effect was most notable aboard ship, compared to shore-based test sites. Also, a number of features within the HPFD and ePerformance systems were identified as sources of error and

frustration. Nonsupervisors tended to have more difficulty using specific features of the Web-based system, including general navigation and opening documents, cutting and pasting text, and locating buttons needed to perform document functions. Overall, supervisors reported feeling more comfortable using the system.

Secondary issues included concerns about the impact of HPFD and ePerformance systems on the performance appraisal process and on Navy culture. These concerns were most prominently observed among supervisors, who in posttest surveys tended to rate the Web-based performance management system lower than the current EVAL/FITREP system. For the performance appraisal process concerns, users expressed concern for the level within the organization that HPFD and ePerformance systems appraisals can be delegated to and the proposed periodicity of performance counseling and appraisal (i.e., based on report onboard date). In terms of cultural concerns, users raised issues with the amount of text allowed for performance comments within each performance dimension, the method of providing performance feedback, and the need for a thorough yet simple communication/training plan. When discussing the amount of text allowed, usability testing participants were divided between the importance of text for performance appraisal and the unintentional cultural pressure that may be created by the ability to include text for promotion and selection boards. Nearly all participants noted the importance of being assured that the HPFD and ePerformance processes include personal, face-to-face interaction with supervisors and subordinates. This aligns with a Navy culture that encourages strong personal and professional relationships between work-center supervisors and their subordinates. Also, users commented on the importance of a thorough, well-planned communication plan for informing and educating Sailors and commands about the systems Navy-wide.

3 Study Objectives

The objectives for this follow-up study, Usability Study II, were to capture both quantitative and qualitative data from participants to identify potential sources of error and user burden. Specifically, the objectives of this study were to conduct

- usability tests on the HPFD and ePerformance systems to examine the effect of system modifications implemented following Usability Study I on system and function usability by collecting data on the type and frequency of user errors, user reactions to the system, and self-reported user satisfaction with the system;

- usability tests on the HPFD and ePerformance systems to examine the functionality and usability of document workflow within a command or unit by collecting data on the type and frequency of user errors, user reactions to the system, and self-reported user satisfaction with the system; and

- focus group interviews with (1) participants to assess document workflow, cultural concerns, and the impact of system modifications and (2) senior Navy performance management/appraisal stakeholders to explore cultural concerns and identify recommendations to facilitate implementation of a new performance management/appraisal system.

All research instruments and procedures, including participant informed consent forms for both the usability testing and focus group interviews were reviewed and approved by the research team's Institutional Review Board (IRB). Participants were briefed on the purpose of the study and were asked to read and sign an informed consent form and return it to their respective task leaders. Participants were treated in accordance with the guidelines published by the American Psychological Association (American Psychological Association [APA], 2001). No adverse events occurred during the course of this study.

4   Participants

To efficiently collect data, the project manager identified local, on-site liaisons to assist in participant recruiting, scheduling, and study logistics. Instructions sent to the on-site liaisons described the criteria for selecting potential participants—supervisory and nonsupervisory personnel assigned to operational and shore commands or units, ranging in paygrade from E2 through the senior enlisted and officer communities. These staff reflected typical HPFD and ePerformance system users at each of the respective commands involved in usability testing.

Table 1 illustrates the number and types of usability testing participants across the three test sites. Usability Study II (conducted in 2005) participants are compared to Usability Study I (conducted in 2004) participants for reference. The table highlights variations in participants and tasks across all iterations of both years' data collection efforts. Primarily, these variations were caused by (1) design enhancements implemented after early iterations to test as much vertical document flow as possible, (2) site-specific nuances that limited the types of participants and amount of system use, and (3) system failures that limited the tasks participants could complete.

In Usability Study I, not all ePerformance tasks were functional. Emphasis was placed on testing the HPFD system. Hence, test participants in each iteration completed only the HPFD document and the Performance Appraisal 1 document. Nonsupervisors at each of the three sites in Usability Study I completed only the HPFD nonsupervisor version. Supervisors completed the HPFD supervisor version and the Performance Appraisal 1.

In Usability Study II, greater functionality of the ePerformance system had been achieved, and all levels of ePerformance review and approval were tested in addition to the HPFD system. In Iteration 1, participants completed only tasks associated with their specific role in the chain of command. Nonsupervisors completed only the HPFD nonsupervisor version. Supervisors who were assigned to the HPFD task completed only the HPFD supervisor version. Supervisors who were assigned as ePerformance raters or approvers completed only the tasks associated with their specific level. First-level raters completed only Performance Appraisal 1, second-level raters completed only Performance Appraisal 2, and final approvers completed only the Final Approval form.

Table 1. Usability Test Participants—Number of Users and Appraisal Tasks Completed

| Test Site and Year | Location | Number of Appraisal Tasks Completed | | | | | Total Number of Usability Participants/ Test Site |
| | | Nonsupervisors | Supervisors | | | | |
		HPFD-NS	HPFD-S	1LR	2LR	FA	
1—2004	Naval Air Station Brunswick, Brunswick, ME	7	14		—	—	21
1—2005	NAVMETOCCEN, Norfolk, VA	2	2	3	3	3	13
2—2004	USS KITTY HAWK (CV 63),Yokosuka, Japan	5	12		—	—	17
2—2005	USS HOWARD (DDG 83), San Diego, CA	3	—	3	3	3	12
3—2004	Kitsap Naval Base—Bangor, Bangor, WA	9	10		—	—	19
3—2005	BUPERS, Arlington, VA	—	3	2	2	2	9

HPFD-NS = HPFD session for nonsupervisory personnel, HPFD-S = HPFD session for supervisors,
1LR = first-level rater, 2LR = second-level rater, FA = final approver.

After Iteration 1, it became clear that ePerformance system study participants needed more context to understand their tasks, since the performance appraisals and especially the final approval are such short documents. In Iterations 2 and 3, supervisors completed more tasks within the system that were relevant to their workplace role—that is, they completed tasks for all levels up to their level of supervision:

- Lower-level supervisors completed only the HPFD supervisor document.
- Lower mid-level supervisors (first-level raters) completed the HPFD supervisor document and Performance Appraisal 1.
- Upper mid-level supervisors (second-level raters) completed the HPFD supervisor document, Performance Appraisal 1, and Performance Appraisal 2.
- In Iteration 2, senior-level supervisors completed the HPFD supervisor document, Performance Appraisal 1, Performance Appraisal 2, and the Final Approver form. In Iteration 3, senior-level supervisors completed only the Final Approver form and were informed of the other tasks in a personal interview.

No nonsupervisors were available for testing in Iteration 3.

4.1   Iteration 1: Naval Meteorology and Oceanographic Center (NAVMETOCCEN)

Iteration 1 of usability testing took place at NAVMETOCCEN in Norfolk, Virginia, from 25 to 28 July 2005. A total of 13 Navy personnel participated in the usability tests. Of the 13 personnel, 2 were nonsupervisors and 11 were supervisors. Users of the ePerformance system consisted of first- and second-level raters as well as final document approvers. HPFD system users included both supervisors and nonsupervisors.

4.2   Iteration 2: USS HOWARD (DDG 83)

Iteration 2 took place onboard USS HOWARD (DDG 83) in San Diego, California, from 29 to 31 August 2005. A total of 12 Navy personnel participated in the usability testing. Users of the ePerformance system consisted of first- and second-level raters as well as final document approvers. HPFD system users included both supervisors and nonsupervisors. For this usability testing, participants who served as ePerformance system first-level raters, second-level raters, and final approvers also completed the HPFD supervisory document. The two sets of tasks were combined for the second and third iterations because we learned in Iteration 1 at NAVMETOCCEN that Sailors needed a better understanding of the whole system to be able to make comments and suggestions for improvement. Usability testing participants were scheduled into roles that reflected their current roles in the performance appraisal process (e.g., a Leading Petty Officer filled the role of a work center supervisor/first-level rater, Leading Chief Petty Officers served as first- and second-level raters, and Division Officers served as first- and second-level raters and final approvers).

4.3   Iteration 3: Bureau of Naval Personnel (BUPERS)

Iteration 3 took place at the Bureau of Naval Personnel in Arlington, Virginia, from 14 to 16 November 2005. A total of 12 Navy personnel participated in the usability tests, but only nine tests generated complete usability and survey data suitable for analysis due to various scheduling conflicts. Users of the ePerformance system at BUPERS consisted of first- and second-level raters as well as final document approvers. HPFD system users were supervisors and completed only supervisory HPFD sessions. No nonsupervisors were involved in the testing at BUPERS. Usability testing participants were scheduled into roles that reflected their current roles in the

performance appraisal process. For this round of testing, senior enlisted and officers (i.e., Master Chiefs, Lieutenant Commanders, and Commanders) served as work center supervisor/first-level raters, mid-grade to senior officers (i.e., Lieutenant Commanders, Commanders, and Captains) served as first- and second-level raters, and senior officers and government civilians (i.e., Captains, Rear Admirals [Lower Half, Upper Half], and Senior Executive Service civilian personnel) served as final approvers.

5 Instruments and Procedures

5.1   Usability Scenarios

Usability scenarios were developed to collect critical information on Sailors' experiences using NMCI, NSIPS, and the PeopleSoft Web-based system screen layouts, performance appraisal item structures, and on-screen features for the Navy's HPFD and ePerformance systems. Additionally, the scenarios were designed to evaluate the effectiveness of the flow of documents through the two systems. The scenarios simulated key work flow tasks for supervisors and nonsupervisors, including completing the HPFD and ePerformance documents, forwarding documents to a supervisor for review and approval, forwarding documents to a final approver for acceptance, and finalizing the evaluation process. Specifically, usability testing protocols and scenarios targeted the following potential problems:

- how Sailors navigated the Web-based NSIPS and PeopleSoft systems,
- time burden associated with using the software to complete specific tasks,
- technical bugs within the test versions of the HPFD and ePerformance documents,
- Sailors' emotive reactions to using the system (such as relaxation, stress, or frustration),
- how Sailors used help text and other supportive documentation such as a Quick Reference Guide,
- adequacy of the Internet connection and NMCI processing speed, and
- ease with which Sailors were able to finalize and route documents through the system.

In an effort to test the HPFD and ePerformance systems in the field, this research study used a portable usability lab—a coordinated system of digital audio and video data capture equipment. The portable usability lab featured professional-grade video monitoring and recording capabilities, including two high-resolution video cameras with silent remote control pan, tilt, zoom, and focus. This equipment enabled the usability team to videotape for later analysis each participant's on-screen activity, including clicking on various links and entering text data, as well as their facial expressions.

Following the best practices in usability testing described previously, an iterative approach with three separate rounds of usability testing was used. To obtain the perspectives and assess the experiences of the diverse Navy workforce, it was important to include participants from a variety of work environments in different geographic locations. As a result, the current research plan included usability testing among Sailors in a variety of communities (i.e., naval meteorology, surface, and personnel management) aboard ship (i.e., USS HOWARD [DDG 83]) and shore-based (i.e., NAVMETOCCEN and BUPERS.)

5.2   Usability Survey

Two paper-and-pencil self-administered surveys—pretest and posttest surveys—were developed to obtain Navy personnel's subjective impressions of the HPFD and ePerformance systems. The objective of the participant surveys was to obtain data on users' subjective reactions to the Web-based tool and assess ease of use, professional value, personal value, and overall satisfaction with the Navy's new performance appraisal and management tool. The surveys were designed to replicate measures collected in Usability Study I for consistency of analyses across time.

The pretest survey included items related to participant demographics (e.g., age, gender, race/ethnicity, education, paygrade, and time on active duty), frequency of computer use both at home and at work, prior experience with PeopleSoft software, satisfaction with the current performance appraisal process, satisfaction with the advancement/promotion process, and perceived difficulty with the HPFD and ePerformance systems prior to use. Items assessing satisfaction with the current performance appraisal process and satisfaction with the advancement/promotion process were adapted from the 2000 Navy-wide Personnel Survey (Olmsted & Underhill, 2003).

The posttest survey asked participants to report their perceptions about completing the tasks in the usability testing portion of this study. Specifically, items asked about perceived comfort in completing the tasks, how successful they believed they were in completing the tasks, ease of use compared to other systems, overall perceived ease of use, how difficult the system was to understand, perceived appearance of the system, perceived efficiency of the system, acclimation or gradual improvement of use while using the system, satisfaction with the current

performance appraisal process, satisfaction with the advancement/promotion process, and overall satisfaction with the pilot HPFD and ePerformance systems.

5.3   Analysis Methods

Analysis of data repeated the same models from Usability Study I, for purposes of consistency. Independent variables were supervisor status (supervisor/nonsupervisor) and test site location (NAVMETOCCEN, USS HOWARD (DDG 83), or BUPERS). Dependent variables were task durations, coded usability errors, and usability survey responses. These replicated analyses were enhanced with an additional set of analyses—comparisons of the differences between HPFD results from Usability Study I and Usability Study II. (Since Usability Study I did not include the extensive ePerformance workflow testing that Usability Study II operationalized, a comparison of ePerformance results between the two usability studies is not available.) These additional analyses allow the researchers to gauge the effect of HPFD system improvements that were recommended based on Usability Study I findings.

Data from usability tests and surveys were analyzed with three analytic techniques: t-tests, analysis of variance (ANOVA), and Fisher's exact test. T-tests were used when only two groups were involved and the data were continuous (or could be treated as such for testing purposes). A Bonferroni multiple comparison procedure from ANOVA was used for continuous variables involving more than two groups, such as comparing task time and error frequency between locations. If the model was significant, the comparisons were checked for significance at the .05 level. For categorical independent variables, we used the Fisher's exact test from a contingency table. The Fisher's exact test is often used when sample sizes or cell sizes are too small to use a chi-square test. Although the t-test and ANOVA require random, normally distributed samples, cautiously applying these statistical tests to convenience samples is a common practice in the usability testing literature (e.g., Westerman, 1997; Wiedenbeck, 1999; Norman et al., 2000).

In usability testing, researchers typically experimentally manipulate stimuli to compare the effect on system usability between groups or between conditions. While this may be a subject of study in a follow-up full-scale pilot study, the objective of this study was to examine system usability in a group of potential system users. As a result, no experimental effects were

examined, rather, usability was examined between user groups (i.e., supervisors and nonsupervisors and users at different geographic locations).

Given these two constraints, the interpretation of the results should take into account the following points. First, the findings may not be generalizable to the entire Navy population. Generalization may be possible only through large-scale studies employing probability samples of the target population. Second, since this study did not have experimental and control conditions, the correlations between the independent and dependent variables should not necessary be viewed as causal.

6 Results

6.1 Participant Demographics

Demographic data from the user surveys administered in both Usability Study I and Usability Study II appear in Table 2. A majority of study participants in both studies were between the ages of 18 and 44, were enlisted Sailors in the paygrades of E2-E9, over 80 percent of the participants were male, and approximately three-quarters of participants were White. Participants from Usability Study I more frequently had college or graduate degrees and had more Navy experience than participants in Usability Study II.

Table 2. Demographic Characteristics of Usability Participants

	Usability I		Usability II	
	Percent	*N*	Percent	*N*
Age				
18–24	21.05	12	8.82	3
25–34	43.86	25	44.12	15
35–44	31.58	18	26.47	9
45–54	3.51	2	17.65	6
55+	0	0	2.94	1
Paygrade				
E2–E3	12.28	7	0.00	0
E4–E6	36.84	21	35.29	12
E7–E9	28.07	16	26.47	9
W1–W4	0	0	2.94	1
O1–O3	12.28	7	11.76	4
O4 and above	10.53	6	20.59	7
Gender				
Male	89.54	51	82.35	28
Female	10.46	6	17.65	6
Hispanic Ethnicity				
Hispanic	7.02	4	17.65	6
Not Hispanic	92.98	53	79.41	27
Ethnic Origin[a]				
American Indian or Alaska Native	5.26	3	0.00	0
Asian	1.75	1	2.94	1
Black or African American	10.53	6	5.88	2
Native Hawaiian or other Pacific Islander	0	0	0.00	0
White	75.44	43	70.59	24
Other	8.77	5	17.65	6
Educational Background				
Less than high school completed/no diploma	0	0	0.00	0
Alternate degree, GED, home study, or adult-school certification	5.36	3	0.00	0
High school graduate/diploma	23.63	15	8.82	3
Some college, no degree	33.33	19	26.47	9
Associate's degree or other 2-year degree	1.53	6	8.82	3
Bachelor's degree	21.05	12	32.35	11
Master's degree	1.75	1	23.53	8
Doctoral or professional degree	1.75	1	0.00	0
Number of Years in the Navy				
Less than 1 year	3.51	2	0.00	0
1–5 years	31.58	18	14.71	5
5–10 years	8.77	5	14.71	5
10–15 years	28.07	16	26.47	9
More than 15 years	28.07	16	44.12	15

[a] Note that the total count for "Ethnic Origin" for Usability I is 58, one more than the number of participants who completed the usability survey. For Usability II the total is 33, one less than the total study population because one participant did not provide a response for this question. Ethnic origin is a "mark all that apply" question and some users selected multiple answers.

6.2   Task Durations

An examination of the average completion times required for each task in Usability Study II provides initial information on the relative demands placed on the users, comparing supervisors and nonsupervisors and users at the three locations. Longer average completion times may indicate increased burden. Table 3 displays estimates of the average completion time (presented in seconds) for each HPFD usability task, as well as the results of significance tests of the differences between groups.

None of the HPFD tasks yielded statistically significant differences across groups in terms of average completion time. Looking only at the overall completion time for each task, it is apparent that Task 1—Complete the CBT Tutorial for Nonsupervisors, Task 4—Complete the HPFD, and Task 10—Enter a performance note, took the longest to complete. This pattern, observed in Usability Study I, is not surprising given that these tasks are designed to be more intensive with more steps required than something simple like opening a document or checking spelling.

Table 3.   Average Time to Complete HPFD Usability Testing Task by Task, Supervisor Status, and Location

				Supervisor Status				Location				
	Overall		Supervisor (S)		Nonsup. (NS)		Norfolk (N)		USS HOWARD (H)		BUPERS (B)	
Task Description	(n)	Avg. Time*	(n)	Avg. Time*	(n)	Avg. Time*	(n)	Avg. Time*	(n)	Avg. Time*	(n)	Avg. Time*
Task 1: Complete the CBT Tutorial for Nonsupervisors	4	564.5	3	712.0	1	122.0	1	122.0	—	—	3	712.0
Task 2: Log in to NSIPS	32	68.1	27	70.4	5	56.0	12	70.8	11	86.4	9	42.2
Task 3: Locate and open the HPFD	24	88.9	19	91.4	5	79.6	4	101.0	12	91.3	8	79.4
Task 4: Complete the HPFD	23	645.5	18	633.2	5	690.0	4	502.3	12	654.3	7	712.3
Task 5: Check spelling for three sections of the HPFD	21	71.6	16	63.1	5	99.0	4	60.5	11	72.5	6	77.5
Task 6: Check target behaviors for one dimension of the HPFD	21	49.2	16	46.9	5	56.6	4	60.5	12	52.3	5	32.6
Task 7: Cut and paste within the HPFD	15	63.1	10	58.0	5	73.2	4	74.8	7	55.4	4	64.8
Task 8: Collapse all sections of the HPFD	19	24.1	14	23.1	5	27.0	4	18.3	12	24.1	3	32.0
Task 9: Submit the HPFD to your Supervisor	23	70.5	18	58.0	5	115.4	4	49.8	12	88.7	7	51.1
Task 10: Enter a performance note	15	129.7	10	124.5	5	140.0	4	120.0	8	132.0	3	136.3

Note: The Bonferroni t-test was used to account for multiple comparisons involving the location variable. Variances of the different levels of a subgroup have been assumed to be equal even in situations when it could not be confirmed from the data because of sparseness. No results were significant at the .05 level.

*Time was measured in seconds.

Table 4 compares the overall results of Usability Study I and Usability Study II in terms of estimates of average time to complete tasks. Overall, task completion times were shorter for Usability Study II than for Usability Study I. Only Task 9—Submit the HPFD to your Supervisor, took more time in the second phase of testing, but this difference was not statistically significant. Only two tasks yielded significantly different times. Task 2—Log in to NSIPS, was completed in 68.1 seconds on average in Usability Study II compared to 323.8 seconds on average in Usability Study I. Task 3—Locate and open the HPFD, was completed in 88.9 seconds on average in Usability Study II compared to 212.0 seconds on average in Usability Study I.

Table 4. **Average Time to Complete HPFD Usability Testing Task by Task and Study Period**

| | | | Study Period | | | |
| | Overall | | Usability I | | Usability II | |
Task Description	(n)	Avg. Time*	(n)	Avg. Time*	(n)	Avg. Time*
Task 1: Complete the CBT Tutorial for Nonsupervisors	32	1,517.3	26	1,633.6	6	1,013.5
Task 2: Log in to NSIPS	70	206.9	38	323.8^{U2}	32	68.1^{U1}
Task 3: Locate and open the HPFD	69	169.2	45	212.0^{U2}	24	88.9^{U1}
Task 4: Complete the HPFD	74	694.3	51	716.3	23	645.5
Task 5: Check spelling for three sections of the HPFD	59	82.2	38	88.1	21	71.6
Task 6: Check target behaviors for one dimension of the HPFD	57	49.7	36	50.1	21	49.2
Task 7: Cut and paste within the HPFD	50	81.4	35	89.3	15	63.1
Task 8: Collapse all sections of the HPFD	60	29.6	41	32.1	19	24.1
Task 9: Submit the HPFD to your Supervisor	61	66.3	38	63.7	23	70.5
Task 10: Enter a performance note	54	139.6	39	143.5	15	129.7

[U1] *significantly different from Usability Study I at the .05 level (t-test).*
[U2] *significantly different from Usability Study II at the .05 level (t-test).*
Time was measured in seconds.

Table 5 displays the average completion times for Tasks 11 through 41 in Usability Study II only. Tasks 11 through 41 comprised the ePerformance system component of the usability testing (ePerformance was not thoroughly tested in Usability Study I). Since only supervisors completed ePerformance scenarios, this table shows only the overall results and the differences

between the three locations. Only two tasks yielded statistically significant differences when compared across three locations:

- For Task 13—Create a Performance Appraisal 1, Sailors at NAVMETOCCEN had statistically significant longer durations on average (285.0 seconds) than Sailors at BUPERS (99.0 seconds).
- For Task 22—Finalize and send to the reviewer of Performance Appraisal 1, Sailors aboard USS HOWARD (DDG 83) completed the task with a statistically significant shorter duration (73.4 seconds) than Sailors at NAVMETOCCEN (121.0 seconds).

Additionally, when only two locations had data for a particular task, four tasks resulted in statistically significant results:

- Task 21—Enter a comment for the reviewer of Performance Appraisal 1, was completed much faster by Sailors aboard USS HOWARD (DDG 83) (25.9 seconds) than by Sailors at NAVMETOCCEN (254.5 seconds).
- Task 27—Check a rating description for Performance Appraisal 2, was completed more quickly by Sailors aboard USS HOWARD (DDG 83) (26.3 seconds) than by Sailors at NAVMETOCCEN (75.7 seconds).
- Task 35—Choose whether you agree/disagree with the Final Performance Appraisal, was completed much more quickly at BUPERS (107.0 seconds) than at NAVMETOCCEN (480.0 seconds).
- Task 37—Return to the Main PeopleSoft Menu, was completed faster aboard USS HOWARD (DDG 83) (7.1 seconds) than at BUPERS (182.0 seconds).

Overall, it appears that, tasks were completed faster aboard USS HOWARD (DDG 83).

Table 5.   Average Time to Complete Usability Testing for ePerformance Tasks 11–41 in Usability Study II, by Location

Task Description	Overall		Location					
			Norfolk (N)		USS HOWARD (H)		BUPERS (B)	
	(n)	Avg. Time*	(n)	Avg. Time*	(n)	Avg. Time*	(n)	Avg. Time*
Task 11: Complete the CBT Tutorial for Supervisors	2	1,911.5	2	1,911.5	—	—	—	—
Task 12: Log out of PeopleSoft	7	37.3	6	42.8	1	4.0	—	—
Task 13: Create a Performance Appraisal 1	17	168.9	3	285.0^B	9	169.0	5	99.0^N
Task 14: Locate and open the Performance Appraisal 1	17	128.4	3	135.3	9	106.1	5	164.4
Task 15: Complete the Performance Appraisal 1	15	402.4	3	691.3	9	313.7	3	379.7
Task 16: Save the work you've done in Performance Appraisal 1	8	71.8	3	45.0	5	87.8	—	—
Task 17: Check a rating description for Performance Appraisal 1	8	40.1	3	36.7	5	42.2	—	—
Task 18: Check spelling for two sections of Performance Appraisal 1	5	58.0	2	116.0	3	19.3	—	—
Task 19: Check language for Performance Appraisal 1	11	31.8	3	70.3	6	20.0	2	9.5
Task 20: Calculate ratings for Performance Appraisal 1	11	9.6	3	11.3	5	8.0	3	10.7
Task 21: Enter a comment for the reviewer of Performance Appraisal 1	9	76.7	2	254.5^H	7	25.9^N	—	—
Task 22: Finalize and send to the reviewer of Performance Appraisal 1	17	115.9	3	121.0	9	73.4^B	5	189.4^H
Task 23: Check e-mail account for Performance Appraisal message prompt	3	0.0	3	0.0	—	—	—	—
Task 24: Locate and open the Performance Appraisal 2	11	131.5	3	216.7	6	98.0	2	104.0
Task 25: Complete the Performance Appraisal 2	11	383.5	3	557.3	6	324.2	2	301.5
Task 26: Save the work you've done for Performance Appraisal 2	5	61.2	3	59.7	2	63.5	—	—

(continued)

Table 5. Average Time to Complete Usability Testing for ePerformance Tasks 11–41 in Usability Study II, by Location (continued)

Task Description	Overall		Location					
			Norfolk (N)		USS HOWARD (H)		BUPERS (B)	
	(n)	Avg. Time*	(n)	Avg. Time*	(n)	Avg. Time*	(n)	Avg. Time*
Task 27: Check a rating description for Performance Appraisal 2	6	51.0	3	75.7^H	3	26.3^N	—	—
Task 28: Check spelling for two sections of Performance Appraisal 2	4	14.3	3	17.3	1	5.0	—	—
Task 29: Check language for Performance Appraisal 2	6	16.8	3	15.7	3	18.0	—	—
Task 30: Calculate ratings for Performance Appraisal 2	9	12.8	3	19.7	5	10.2	1	5.0
Task 31: Enter a comment for the final approver of Performance Appraisal 2	9	57.6	3	53.3	5	70.2	1	7.0
Task 32: Finalize and send to the final approver of Performance Appraisal 2	9	150.8	3	129.7	5	175.0	1	93.0
Task 33: Locate and open the Final Performance Appraisal awaiting your review and approval	6	205.7	2	239.5	1	181.0	3	191.3
Task 34: Review any comments made by reviewers in the Final Performance Appraisal	5	167.0	2	78.5	—	—	3	226.0
Task 35: Choose whether you agree/disagree with the Final Performance Appraisal	5	256.2	2	480.0^B	—	—	3	107.0^N
Task 37: Return to the Main PeopleSoft Menu	7	32.1	—	—	6	7.1^B	1	182.0^H
Task 39: Locate and open the Employee Reviewed Final Performance Appraisal	1	63.0	1	63.0	—	—	—	—
Task 40: Review and enter final comments into the Employee Reviewed Final Performance Appraisal	1	215.0	1	215.0	—	—	—	—
Task 41: Finalize and complete the Employee Reviewed Final Performance Appraisal	2	130.0	2	130.0	—	—	—	—

Note: The Bonferroni t-test was used to account for multiple comparisons involving the location variable. Variances of the different levels of a subgroup have been assumed to be equal even in situations when it could not be confirmed from the data because of sparseness.^{B} significantly different from BUPERS at the .05 level (Bonferroni t-test).

^{H} *significantly different from USS HOWARD at the .05 level (Bonferroni t-test).*

^{N} *significantly different from Norfolk (NAVMETOCCEN) at the .05 level (Bonferroni t-test).*

*Time was measured in seconds.

6.3. Usability Errors

Usability errors are presented according to three different dimensions: total error frequency per task, rate of error occurrence per task, and the most frequent error category per task. Each of these analyses yields a different perspective on the types of usability problems users experienced. They enhance simple duration analyses by revealing the specific causes of usability problems that resulted in long task durations.

6.3.1 Total Error Frequency

Table 6 presents the estimates of the *total error frequency*—that is, the total number of errors across all types of error for the 10 HPFD tasks. (The coding scheme used to identify errors can be viewed in Appendix A.) Total error frequency varied from task to task because time requirements and task complexity varied. Only Task 1—Complete the CBT Tutorial for Nonsupervisors, had a statistically significant difference in the total error frequency across locations, with only four observations and one error overall in the particular task. Overall, the greatest error frequencies occurred in Task 3—Locate and open the HPFD, and Task 4—Complete the HPFD. These are two central tasks of completing the HPFD form. Based on the durations analysis, it is not surprising that Task 4—Complete the HPFD, yielded a high total error frequency, simply due to the fact that test participants spent more time completing this task than other tasks.

Table 6. Total Error Frequency for HPFD by Task, Supervisor Status, and Location

Task Description[a]	Overall		Supervisor Status				Location					
			Supervisor (S)		Nonsup. (NS)		Norfolk (N)		USS HOWARD (H)		BUPERS (B)	
	(n)	Avg. # Errors	(n)	Avg. # Errors	(n)	Avg. # Errors	(n)	Avg. # Errors	(n)	Avg. # Errors	(n)	Avg. # Errors
Task 1: Complete the CBT Tutorial for Nonsupervisors	4	0.25	3	0.00	1	1.00	1	1.00^B	—	—	3	0.00^N
Task 2: Log in to NSIPS	32	0.50	27	0.48	5	0.60	12	1.25	11	0.09	9	0.00
Task 3: Locate and open the HPFD	24	5.79	19	6.00	5	5.00	4	3.50	12	6.58	8	5.75
Task 4: Complete the HPFD	23	3.04	18	3.06	5	3.00	4	2.00	12	4.17	7	1.71
Task 5: Check spelling for three sections of the HPFD	21	0.90	16	0.81	5	1.20	4	0.50	11	1.09	6	0.83
Task 6: Check target behaviors for one dimension of the HPFD	21	1.24	16	1.19	5	1.40	4	1.50	12	1.42	5	0.60
Task 7: Cut and paste within the HPFD	15	0.47	10	0.40	5	0.60	4	0.25	7	0.57	4	0.50
Task 8: Collapse all sections of the HPFD	19	0.47	14	0.57	5	0.20	4	0.25	12	0.42	3	1.00
Task 9: Submit the HPFD to your Supervisor	23	1.65	18	1.28	5	3.00	4	1.00	12	2.33	7	0.86
Task 10: Enter a performance note	15	1.93	10	1.90	5	2.00	4	2.25	8	2.00	3	1.33

Note: The Bonferroni t-test was used to account for multiple comparisons involving the location variable. Variances of the different levels of a subgroup have been assumed to be equal even in situations when it could not be confirmed from the data because of sparseness.

[a] *Tasks 11 through 41 were completed only by study subjects with supervisor status. Therefore, these tasks were excluded from the analysis.*

[B] *significantly different from BUPERS at the .05 level (Bonferroni t-test).*

[N] *significantly different from Norfolk (NAVMETOCCEN) at the .05 level (Bonferroni t-test).*

Compared to Usability Study I, like the duration analysis, the total error frequency analysis yielded fewer statistically significant differences across groups. In Usability Study I, two tasks, Task 1—Complete the CBT Tutorial for Nonsupervisors, and Task 4—Complete the HPFD, had statistically significant differences among locations. Table 7 shows the results of the full comparison of the overall total error frequency from Usability Study I to Usability Study II. A pattern of reduced total error frequency does not hold for all tasks. Five tasks actually experienced an increase in total error frequency, although the differences were small and not statistically significant. However, the four tasks with statistically significant differences from Usability I to Usability II all show decreases in total error frequency.

- Usability Study II participants exhibited an average of 0.33 errors in Task 1— Complete the CBT Tutorial for Nonsupervisors, compared to an average frequency of 21.38 errors in Usability Study I. Note that only six tutorials were completed in Usability Study II, compared to 26 in Usability Study I.

- When completing Task 2—Log in to NSIPS, Usability Study II participants experienced an average of 0.50 errors compared to 1.45 errors in Usability Study I. Since logging in to NSIPS was required significantly less time to complete in Usability Study II, the decrease in total error frequency suggests that the system showed improvement.

- The total error frequency of Task 5—Check spelling for three sections of the HPFD, was reduced from 1.89 in Usability Study I to 0.90 in Usability Study II.

- Usability Study II participants also had fewer errors completing Task 8—Collapse all sections of the HPFD. The total error frequency for Usability Study II was 0.47 compared to 1.22 in Usability Study I.

Overall, the comparison of total error frequency between Usability Study I and Usability Study II finds small but significant changes in the expected direction of reduced errors.

Table 7.   Total Error Frequency for HPFD by Task and Study Period

			Study Period			
	Overall		Usability I		Usability II	
Task Description	(n)	Avg. # Errors[a]	(n)	Avg. # Errors[a]	(n)	Avg. # Errors[a]
Task 1: Complete the CBT Tutorial for Nonsupervisors	32	17.44	26	21.38[U2]	6	0.33[U1]
Task 2: Log in to NSIPS	70	1.01	38	1.45[U2]	32	0.50[U1]
Task 3: Locate and open the HPFD	69	5.06	45	4.67	24	5.79
Task 4: Complete the HPFD	74	2.55	51	2.33	23	3.04
Task 5: Check spelling for three sections of the HPFD	59	1.54	38	1.89[U2]	21	0.90[U1]
Task 6: Check target behaviors for one dimension of the HPFD	57	1.16	36	1.11	21	1.24
Task 7: Cut and paste within the HPFD	50	0.62	35	0.69	15	0.47
Task 8: Collapse all sections of the HPFD	60	0.98	41	1.22[U2]	19	0.47[U1]
Task 9: Submit the HPFD to your Supervisor	61	1.57	38	1.53	23	1.65
Task 10: Enter a performance note	54	1.87	39	1.85	15	1.93

[a] *Two types of errors that were tracked in Usability Study I were not tracked in Usability Study II. Therefore, those two types of errors were not counted in these combined study analyses. Consequently, for some tasks, averages for Usability Study I may appear smaller than on previous tables.*
[U1] *significantly different from Usability Study I at the .05 level (t-test)*
[U2] *significantly different from Usability Study II at the .05 level (t-test)*

Table 8 displays the average total error frequency for Usability Study II Tasks 11 through 41, the ePerformance tasks. Only Task 19—Check language for Performance Appraisal 1, had significantly different outcomes when all three locations were tested. Participants aboard USS HOWARD (DDG 83) averaged 0.17 errors in Task 19, whereas participants at NAVMETOCCEN averaged 1.67 errors. BUPERS did not yield a significantly different outcome for Task 19 when compared to the other three locations. Two tasks resulted in statistically significant differences when only two locations were compared. Task 18—Check spelling for two sections of Performance Appraisal 1, resulted in no errors aboard USS HOWARD (DDG 83) and an average of 2.5 errors at NAVMETOCCEN. Similarly, Task 21— Enter a comment for the reviewer of Performance Appraisal 1, resulted in no errors aboard USS HOWARD (DDG 83) and an average of 1.5 errors at NAVMETOCCEN. Overall, it appears that tasks related to creating, locating, and opening Performance Appraisal 1 and Performance Appraisal 2 generated the most errors across locations. These tasks require the user to navigate through the complicated series of folders and menus to find the correct document. In general, participants aboard USS HOWARD (DDG 83) experienced fewer errors on average.

Table 8.   Total Error Frequency for ePerformance Tasks 11 through 41 in Usability Study II, by Task and Location

| | | | Location | | | | | |
| | Overall | | Norfolk (N) | | USS HOWARD (H) | | BUPERS (B) | |
Task Description	(n)	Avg. # Errors	(n)	Avg. # Errors	(n)	Avg. # Errors	(n)	Avg. # Errors
Task 11: Complete the CBT Tutorial for Supervisors	2	0.50	2	0.50	—	—	—	—
Task 12: Log out of PeopleSoft	7	0.14	6	0.17	1	0.00	—	—
Task 13: Create a Performance Appraisal 1	17	9.65	3	15.00	9	9.22	5	7.20
Task 14: Locate and open the Performance Appraisal 1	17	3.35	3	4.00	9	3.56	5	2.60
Task 15: Complete the Performance Appraisal 1	15	1.60	3	6.33	9	0.22	3	1.00
Task 16: Save the work you've done in the Performance Appraisal 1	8	0.88	3	1.67	5	0.40	—	—
Task 17: Check a rating description for Performance Appraisal 1	8	1.50	3	2.67	5	0.80	—	—
Task 18: Check spelling for two sections of Performance Appraisal 1	5	1.00	2	2.50^H	3	0.00^N	—	—
Task 19: Check language for Performance Appraisal 1	11	0.64	3	1.67^H	6	0.17^N	2	0.50
Task 20: Calculate ratings for Performance Appraisal 1	11	0.27	3	1.00	5	0.00	3	0.00
Task 21: Enter a comment for the reviewer of Performance Appraisal 1	9	0.33	2	1.50^H	7	0.00^N	—	—
Task 22: Finalize and send to the reviewer of Performance Appraisal 1	17	1.29	3	2.33	9	1.22	5	0.80
Task 23: Check e-mail account for Performance Appraisal message prompt	3	0.00	3	0.00	—	—	—	—
Task 24: Locate and Open the Performance Appraisal 2	11	5.55	3	3.67	6	5.83	2	7.50
Task 25: Complete the Performance Appraisal 2	11	1.55	3	3.00	6	1.33	2	0.00
Task 26: Save the work you've done for Performance Appraisal 2	5	1.00	3	1.00	2	1.00	—	—

(continued)

Table 8. Total Error Frequency for ePerformance Tasks 11 through 41 in Usability II, by Task and Location (continued)

Task Description	Overall		Norfolk (N)		USS HOWARD (H)		BUPERS (B)	
	(n)	Avg. # Errors	(n)	Avg. # Errors	(n)	Avg. # Errors	(n)	Avg. # Errors
Task 27: Check a rating description for Performance Appraisal 2	6	0.50	3	0.67	3	0.33	—	—
Task 28: Check spelling for two sections of Performance Appraisal 2	4	0.75	3	1.00	1	0.00	—	—
Task 29: Check language for Performance Appraisal 2	6	0.17	3	0.00	3	0.33	—	—
Task 30: Calculate ratings for Performance Appraisal 2	9	0.11	3	0.33	5	0.00	1	0.00
Task 31: Enter a comment for the final approver of Performance Appraisal 2	9	0.11	3	0.00	5	0.20	1	0.00
Task 32: Finalize and send to the final approver of Performance Appraisal 2	9	1.11	3	0.67	5	1.20	1	2.00
Task 33: Locate and open the Final Performance Appraisal awaiting your review and approval	6	3.33	2	3.00	1	5.00	3	3.00
Task 34: Review any comments made by reviewers in the Final Performance Appraisal	5	0.60	2	0.50	—	—	3	0.67
Task 35: Choose whether you agree/disagree with the Final Performance Appraisal	5	1.20	2	2.00	—	—	3	0.67
Task 37: Return to the Main PeopleSoft Menu	7	0.00	—	—	6	0.00	1	0.00
Task 39: Locate and open the Employee Reviewed Final Performance Appraisal	1	0.00	1	0.00	—	—	—	—
Task 40: Review and enter final comments into the Employee Reviewed Final Performance Appraisal	1	1.00	1	1.00	—	—	—	—
Task 41: Finalize and complete the Employee Reviewed Final Performance Appraisal	2	0.50	2	0.50	—	—	—	—

Note: The Bonferroni t-test was used to account for multiple comparisons involving the location variable. Variances of the different levels of a subgroup have been assumed to be equal even in situations when it could not be confirmed from the data because of sparseness.
[H] *significantly different from USS HOWARD at the .05 level (Bonferroni t-test).*
[N] *significantly different from Norfolk (NAVMETOCCEN) at the .05 level (Bonferroni t-test).*

6.3.2 Rate of Error Occurrence

Table 9 presents the *rate of error occurrence* for each HPFD task in Usability Study II. The rate of error occurrence is the percentage of total cases (i.e., total number of usability tests conducted for each task) in which errors occurred. The rate of error occurrence may be a better measure of usability problems than total error frequency because it indicates recurring usability errors for a given task as opposed to the total number of errors, which could be skewed by a particularly problematic case. According to the overall results presented in Table 9, errors occurred 93.3 percent of the time in Task 10—Enter a performance note, and 91.7 percent of the time in Task 3—Locate and open the HPFD. The tasks with the next highest rates of error of occurrence were

- Task 4—Complete the HPFD, 78.3 percent rate of error occurrence;
- Task 9—Submit the HPFD to your Supervisor, 60.9 percent rate of error occurrence;
- Task 5—Check spelling for three sections of the HPFD, 47.6 percent rate of error occurrence; and
- Task 6—Check target behaviors for one dimension of the HPFD, also 47.6 percent rate of error occurrence.

None of the comparisons across groups, defined by supervisor status or location, generated statistically significant differences in the rate of error occurrence.

Table 9. Rate of Error Occurrence for HPFD by Task, Supervisor Status, and Location

Task Description	Overall (n)	Overall Avg. % Errors	Supervisor (S) (n)	Supervisor (S) Avg. % Errors	Nonsup. (NS) (n)	Nonsup. (NS) Avg. % Errors	Norfolk (N) (n)	Norfolk (N) Avg. % Errors	USS HOWARD (H) (n)	USS HOWARD (H) Avg. % Errors	BUPERS (B) (n)	BUPERS (B) Avg. % Errors
			Supervisor Status				**Location**					
Task 1: Complete the CBT Tutorial for Nonsupervisors	4	25.0	3	0.0	1	100.0	1	100.0	—	—	3	0.0
Task 2: Log in to NSIPS	32	9.4	27	7.4	5	20.0	12	16.7	11	9.1	9	0.0
Task 3: Locate and open the HPFD	24	91.7	19	94.7	5	80.0	4	75.0	12	100.0	8	87.5
Task 4: Complete the HPFD	23	78.3	18	77.8	5	80.0	4	75.0	12	91.7	7	57.1
Task 5: Check spelling for three sections of the HPFD	21	47.6	16	43.8	5	60.0	4	25.0	11	63.6	6	33.3
Task 6: Check target behaviors for one dimension of the HPFD	21	47.6	16	50.0	5	40.0	4	75.0	12	41.7	5	40.0
Task 7: Cut and paste within the HPFD	15	40.0	10	40.0	5	40.0	4	25.0	7	42.9	4	50.0
Task 8: Collapse all sections of the HPFD	19	31.6	14	35.7	5	20.0	4	25.0	12	33.3	3	33.3
Task 9: Submit the HPFD to your Supervisor	23	60.9	18	61.1	5	60.0	4	50.0	12	75.0	7	42.9
Task 10: Enter a performance note	15	93.3	10	90.0	5	100.0	4	75.0	8	100.0	3	100.0

Note: No differences in this table were statistically significant at the .05 level using the Fisher's exact test to test for association between the percentage of error occurrence and the independent variable.

Table 10 presents the results of comparing the rate of error occurrence for each HPFD task from Usability Study I to Usability Study II. As with the results comparing total error frequency, there is not a consistent decrease in the rate of error occurrence between the first study and the current study. However, for the four statistically significant tasks, three show a clear decrease in the rate of error occurrence. One shows a marked increase, a cause of some concern.

- Errors occurred while completing Task 1—Complete the CBT Tutorial for Nonsupervisors, 33.3 percent of the time in Usability Study II compared to 100 percent of the time in Usability Study I.

- Errors occurred while completing Task 2—Log in to NSIPS, only 9.4 percent of the time in Usability Study II, compared to 60.5 percent of the time in Usability Study I.

- Errors occurred while completing Task 5—Check spelling for three sections of the HPFD, 47.6 percent of the time in Usability Study II versus 78.9 percent of the time in Usability Study I.

Table 10. Rate of Error Occurrence for HPFD by Task and Study Period

| | | | Study Period | | | |
| | | | Usability I | | Usability II | |
Task Description	(n)	Avg. % Errors[a] (Overall)	(n)	Avg. % Errors[a]	(n)	Avg. % Errors[a]
Task 1: Complete the CBT Tutorial for Nonsupervisors	32	87.5	26	100.0*	6	33.3*
Task 2: Log in to NSIPS	70	37.1	38	60.5*	32	9.4*
Task 3: Locate and open the HPFD	69	92.8	45	93.3	24	91.7
Task 4: Complete the HPFD	74	68.9	51	64.7	23	78.3
Task 5: Check spelling for three sections of the HPFD	59	67.8	38	78.9*	21	47.6*
Task 6: Check target behaviors for one dimension of the HPFD	57	61.4	36	69.4	21	47.6
Task 7: Cut and paste within the HPFD	50	28.0	35	22.9	15	40.0
Task 8: Collapse all sections of the HPFD	60	50.0	41	58.5	19	31.6
Task 9: Submit the HPFD to your Supervisor	61	68.9	38	73.7	23	60.9
Task 10: Enter a performance note	54	68.5	39	59.0*	15	93.3*

[a] *Two types of errors that were tracked in Usability Study I were not tracked in Usability Study II. Therefore, those two types of errors were not counted in these combined study analyses. Consequently, for some tasks, averages for Usability I may appear smaller than on previous tables.*

* indicates a significant association between the percentage of error occurrence and the independent variable at the .05 level using Fisher's exact test.

- Errors occurred at a greater rate in Task 10—Enter a performance note, increasing from 59.0 percent in Usability Study I to 93.3 percent in Usability Study II.

With the exception of the performance note findings, these results, much like the previous Usability Study I and Usability Study II comparisons, suggest improvement as a result of post-Usability Study I changes. The Performance Note feature was not included in the QRG along with other navigational instructions and function labels, which may account, at least in part, for the higher observed error rate for Task 10—Enter a performance note.

Table 11 displays the rate of error occurrence for the ePerformance tasks in Usability Study II. Two tasks had statistically significant differences in outcomes at different locations. Task 19—Check language for Performance Appraisal 1, errors occurred at a rate of 100 percent at NAVMETOCCEN, 50 percent at BUPERS, and only 16.7 percent aboard USS HOWARD (DDG 83). For Task 21—Enter a Comment for the Reviewer of Performance Appraisal 1, no errors were observed aboard USS HOWARD (DDG 83) but errors occurred at a rate of 100 percent at NAVMETOCCEN. Looking at all users overall, the most error-prone tasks were

- Task 13—Create Performance Appraisal 1, with errors occurring 100 percent of the time;
- Task 14—Locate and Open Performance Appraisal 1, with errors occurring 50 percent of the time;
- Task 24—Locate and Open Performance Appraisal 2, with errors occurring 90.9 percent of the time; and
- Task 32—Finalize and Send to the Reviewer of Performance Appraisal 2, with errors occurring 77.8 percent of the time.

Table 11. Rate of Error Occurrence for ePerformance Tasks 11 through 41 in Usability Study II, by Task and Location

| Task Description | Overall | | Norfolk (N) | | USS HOWARD (H) | | BUPERS (B) | |
				Location				
	(n)	Avg. % Errors	(n)	Avg. % Errors	(n)	Avg. % Errors	(n)	Avg. % Errors
Task 11: Complete the CBT Tutorial for Supervisors	2	50.0	2	50.0	—	—	—	—
Task 12: Log out of PeopleSoft	7	14.3	6	16.7	1	0	—	—
Task 13: Create a Performance Appraisal 1	17	100.0	3	100.0	9	100.0	5	100.0
Task 14: Locate and open the Performance Appraisal 1	17	82.4	3	100.0	9	88.9	5	60.0
Task 15: Complete the Performance Appraisal 1	15	40.0	3	100.0	9	22.2	3	33.3
Task 16: Save the work you've done in the Performance Appraisal 1	8	50.0	3	66.7	5	40.0	—	—
Task 17: Check a rating description for Performance Appraisal 1	8	50.0	3	66.7	5	40.0	—	—
Task 18: Check spelling for two sections of Performance Appraisal 1	5	40.0	2	100.0	3	0.0	—	—
Task 19: Check language for Performance Appraisal 1	11	45.5	3	100.0*	6	16.7*	2	50.0*
Task 20: Calculate ratings for Performance Appraisal 1	11	18.2	3	66.7	5	0.0	3	0.0
Task 21: Enter a comment for the reviewer of Performance Appraisal 1	9	22.2	2	100.0*	7	0.0*	—	—
Task 22: Finalize and send to the reviewer of Performance Appraisal 1	17	70.6	3	100.0	9	66.7	5	60.0
Task 23: Check e-mail account for Performance Appraisal message prompt	3	0.0	3	0.0	—	—	—	—
Task 24: Locate and Open the Performance Appraisal 2	11	90.9	3	100.0	6	83.3	2	100.0
Task 25: Complete the Performance Appraisal 2	11	54.5	3	100.0	6	50.0	2	0.0
Task 26: Save the work you've done for Performance Appraisal 2	5	60.0	3	66.7	2	50.0	—	—

(continued)

Table 11. Rate of Error Occurrence for ePerformanceTasks 11 through 41 in Usability Study II, by Task and Location (continued)

Task Description	Overall (n)	Overall Avg. % Errors	Norfolk (N) (n)	Norfolk (N) Avg. % Errors	USS HOWARD (H) (n)	USS HOWARD (H) Avg. % Errors	BUPERS (B) (n)	BUPERS (B) Avg. % Errors
Task 27: Check a rating description for Performance Appraisal 2	6	50.0	3	66.7	3	33.3	—	—
Task 28: Check spelling for two sections of Performance Appraisal 2	4	50.0	3	66.7	1	0.0	—	—
Task 29: Check language for Performance Appraisal 2	6	16.7	3	0.0	3	33.3	—	—
Task 30: Calculate ratings for Performance Appraisal 2	9	11.1	3	33.3	5	0.0	1	0.0
Task 31: Enter a comment for the final approver of Performance Appraisal 2	9	11.1	3	0.0	5	20.0	1	0.0
Task 32: Finalize and send to the final approver of Performance Appraisal 2	9	77.8	3	66.7	5	80.0	1	100.0
Task 33: Locate and open the Final Performance Appraisal awaiting your review and approval	6	66.7	2	100.0	1	100.0	3	33.3
Task 34: Review any comments made by reviewers in the Final Performance Appraisal	5	40.0	2	50.0	—	—	3	33.3
Task 35: Choose whether you agree/disagree with the Final Performance Appraisal	5	40.0	2	50.0	—	—	3	33.3
Task 37: Return to the Main PeopleSoft Menu	7	0.0	—	—	6	0.0	1	0.0
Task 39: Locate and open the Employee Reviewed Final Performance Appraisal	1	0.0	1	0.0	—	—	—	—
Task 40: Review and enter final comments into the Employee Reviewed Final Performance Appraisal	1	100.0	1	100.0	—	—	—	—
Task 41: Finalize and complete the Employee Reviewed Final Performance Appraisal	2	50.0	2	50.0	—	—	—	—

* indicates a significant association between the percentage of error occurrence and the independent variable at the .05 level from Fisher's exact test.

6.3.3 Most Frequently Occurring Error Type

Table 12 presents the most frequently occurring error category for each HPFD and ePerformance task in Usability Study I and Usability Study II. Statistical testing of errors was not conducted because the most frequently occurring error varied between the two study periods. Testing the variation in errors between years would not withstand the small sample sizes tested. For Usability Study II, the most frequent HPFD "error" overall was referring to the QRG. This was not a true user error within the program *per se*, but was coded as such to capture the frequency with which users relied upon the QRG to complete certain tasks. In 9 of the 10 Usability Study II HPFD tasks, the most frequently occurring error was referring to the QRG. This compares favorably to Usability Study I, when the most frequently occurring error was timing out and when a greater variety of most frequently occurring errors—including navigational errors, button errors, and being unable to set new passwords—were observed.

Table 12 also presents the most frequently occurring ePerformance errors from Usability Study I and Usability Study II. Only a few ePerformance tasks were tested in Usability Study I, so many of the cells in this portion of the table are blank. The most common of the Usability Study II errors was referring to the QRG. Of the 30 ePerformance tasks, referring to the QRG was the most frequently occurring error for 14 unique tasks. The next most common most frequently occurring error was incorrectly clicking on the wrong button—a problem that occurred most frequently in five of the tasks. In Usability Study I, only one error was the most frequently occurring over all five tasks tested: timing out. The reduction in system timeout problems suggests noteworthy functional improvement in the system.

Table 12. Most Frequently Occurring Error by Task and the Error's Average Frequency

Task Description	Usability I (n)	Average Frequency	Error Description	Usability II (n)	Average Frequency	Error Description
HPFD Tasks						
Task 1: Complete the CBT Tutorial for Nonsupervisors	26	17.19	Doesn't follow screen instruction	4	0.25	User retries action
Task 2: Log in to NSIPS	38	2.16	Can't set new passwords	32	0.38	Refers to QRG
Task 3: Locate and open the HPFD	45	2.11	Navigation error	24	5.38	Refers to QRG
Task 4: Complete the HPFD	51	0.65	Refer to info sheet	23	2.52	Refers to QRG
Task 5: Check spelling for three sections of the HPFD	38	0.87	Time out	21	0.67	Refers to QRG
Task 6: Check target behaviors for one dimension of the HPFD	36	0.64	Time out	21	0.90	Refers to QRG
Task 7: Cut and paste within the HPFD	35	0.17	PeopleSoft button error	15	0.27	Refers to QRG
Task 8: Collapse all sections of the HPFD	41	0.63	Time out	19	0.32	Refers to QRG
Task 9: Submit the HPFD to your Supervisor	38	0.39	Time out	23	1.30	Refers to QRG
Task 10: Enter a performance note	39	0.64	Navigation error	15	1.00	Refers to QRG
ePerformance Tasks						
Task 11: Complete the CBT Tutorial for Supervisors	—	—	—	2	0.50	Ask for help
Task 12: Log out of PeopleSoft	—	—	—	7	0.14	Time out
Task 13: Create a Performance Appraisal 1	—	—	—	17	8.47	Refers to QRG
Task 14: Locate and open the Performance Appraisal 1	—	—	—	17	2.88	Refers to QRG
Task 15: Complete the Performance Appraisal 1	—	—	—	15	0.67	Refers to QRG
Task 16: Save the work you've done in the Performance Appraisal 1	—	—	—	8	0.38	Refers to QRG
Task 17: Check a rating description for Performance Appraisal 1	12	0.25	Time out	8	0.88	Refers to QRG
Task 18: Check spelling for two sections of Performance Appraisal 1	11	0.36	Time out	5	0.40	Incorrect click on wrong button
Task 19: Check language for Performance Appraisal 1	18	0.39	Time out	11	0.18	Incorrect click on wrong button
Task 20: Calculate ratings for Performance Appraisal 1	11	0.37	Time out	11	0.09	Failed task
Task 21: Enter a comment for the reviewer of Performance Appraisal 1	—	—	—	9	0.11	Incorrect click on wrong button

(continued)

Table 12. Most Frequently Occurring Error by Task and the Error's Average Frequency (continued)

Task Description	Usability I (n)	Average Frequency	Error Description	Usability II (n)	Average Frequency	Error Description
ePerformance Tasks						
Task 22: Finalize and send to the reviewer of Performance Appraisal 1	7	0.71	Time out	17	0.59	Refers to QRG
Task 23: Check e-mail account for Performance Appraisal message prompt	—	—	—	3	0	(No error occurred)
Task 24: Locate and Open the Performance Appraisal 2	—	—	—	11	4.45	Refers to QRG
Task 25: Complete the Performance Appraisal 2	—	—	—	11	1.27	Refers to QRG
Task 26: Save the work you've done for Performance Appraisal 2	—	—	—	5	0.60	Refers to QRG
Task 27: Check a rating description for Performance Appraisal 2	—	—	—	6	0.33	Refers to QRG
Task 28: Check spelling for two sections of Performance Appraisal 2	—	—	—	4	0.25	Incorrect click on wrong button
Task 29: Check language for Performance Appraisal 2	—	—	—	6	0.17	Refers to QRG
Task 30: Calculate ratings for Performance Appraisal 2	—	—	—	9	0.11	Refers to QRG
Task 31: Enter a comment for the final approver of Performance Appraisal 2	—	—	—	9	0.11	Ask for help
Task 32: Finalize and send to the final approver of Performance Appraisal 2	—	—	—	9	0.67	Refers to QRG
Task 33: Locate and open the Final Performance Appraisal awaiting your review and approval	—	—	—	6	2.17	Refers to QRG
Task 34: Review any comments made by reviewers in the Final Performance Appraisal	—	—	—	5	0.40	Time out
Task 35: Choose whether you agree/disagree with the Final Performance Appraisal	—	—	—	5	0.60	Navigational error
Task 37: Return to the Main PeopleSoft Menu	—	—	—	7	0	(No error occurred)
Task 39: Locate and open the Employee Reviewed Final Performance Appraisal	—	—	—	1	0	(No error occurred)
Task 40: Review and enter final comments into the Employee Reviewed Final Performance Appraisal	—	—	—	1	1.00	Incorrect click on wrong button
Task 41: Finalize and complete the Employee Reviewed Final Performance Appraisal	—	—	—	2	0.50	Time out

Note. ePerformance Tasks (Tasks 11 through 41) were completed only by participants with supervisor status.

6.4   User Ratings

User ratings measured on pretest and posttest surveys were compared between supervisors and nonsupervisors and between test sites using ANOVA. Because of power limitations due to low sample size typical of usability studies, only supervisor versus nonsupervisor analyses appeared to have enough power to explain group differences in the dependent variables. As a result, analyses of user ratings focused on comparisons between supervisory and nonsupervisory user ratings. However, conclusions must be drawn with caution because, due to site logistics, only five nonsupervisors were interviewed in Usability Study II. Tables 13 and 14 present overall average rating scores for questionnaire responses and the average rating scores for supervisors and nonsupervisors, for pretest and posttest surveys.

Questions on the pretest survey asked about the user's satisfaction with the current EVAL/FITREP system and their expectations for using the Web-based system. Results from the analyses of the pretest survey indicated significant differences on only two aspects of Sailors' perceptions of the EVAL/FITREP system. Compared to nonsupervisors, supervisors' ratings were significantly higher for having a clear understanding of the EVAL/FITREP system and on perceptions of fairness in advancement/promotion. The direction and significance of these two variables are consistent with Usability Study I results.

Questions on the posttest survey asked users how they felt about the test version of the Web-based performance management system. Specifically, participants were asked to rate on a scale of 1-5 their perceptions of the Web-based system including their comfort level completing the tasks, ease of using and understanding the system, and how professional and efficient it seemed. Additionally, the same measures of perceptions of the EVAL/FITREP system that were captured in the pretest survey were captured with regard to the Web-based system, including fairness, timeliness, and overall satisfaction. There were no statistically significant differences between supervisors and nonsupervisors.

Table 13. Usability Pretest Survey Outcomes by Supervisor Status

Variable Description	Overall		Supervisor Status			
			Supervisors		Nonsupervisors	
	(n)	Avg. Score	(n)	Avg. Score	(n)	Avg. Score
I have a clear understanding of the present EVAL/FITREP system. 5: Strongly agree ~ 1: Strongly disagree[a]	34	4.35	29	4.45[NS]	5	3.80[S]
My last EVAL/FITREP was fair/accurate. 5: Strongly agree ~ 1: Strongly disagree	34	4.41	29	4.45	5	4.20
My last EVAL/FITREP was conducted in a timely manner. 5: Strongly agree ~ 1: Strongly disagree	34	4.15	29	4.14	5	4.20
I was able to submit my own input at my last EVAL/FITREP. 5: Strongly agree ~ 1: Strongly disagree	34	4.59	29	4.66	5	4.20
My last advancement/promotion recommendation was fair/accurate. 5: Strongly agree ~ 1: Strongly disagree	34	4.53	29	4.62[NS]	5	4.00[S]
I am satisfied with the present Navy EVAL/FITREP system. 5: Strongly agree ~1: Strongly disagree	34	3.41	29	3.41	5	3.40
The most qualified and deserving Sailors score the highest on their EVALs/FITREPs. 5: Strongly agree ~ 1: Strongly disagree	34	3.32	29	3.38	5	3.00
How easy or difficult do you think it will be to use this test version of the performance management system? 5: Very easy ~ 1: Very difficult	34	2.82	29	2.76	5	3.20
How efficient or inefficient do you think the performance management system will be? 5: Very efficient ~ 1: Very inefficient	34	3.47	29	3.48	5	3.40

[a] *The original agreement scale in both survey questionnaires had the opposite endpoints—1 indicated "strongly agree" and 5 "strongly disagree." This scale was reversed in the analysis for convenient presentation.*
[S] *significantly different from supervisors at the .05 level (t-test).*
[NS] *significantly different from nonsupervisors at the .05 level (t-test).*

Table 14. Usability Posttest Survey Outcomes by Supervisor Status

| | Overall | | Supervisor Status | | | |
| | | | Supervisors | | Nonsupervisors | |
Variable Description	(n)	Avg. Score	(n)	Avg. Score	(n)	Avg. Score
How comfortable or uncomfortable did you feel performing the tasks in the test? 5: Very comfortable ~ 1: Very uncomfortable	34	3.74	29	3.72	5	3.80
How certain or uncertain are you that you completed the tasks successfully? 5: Very certain ~ 1: Very uncertain	34	3.74	29	3.66	5	4.20
Compared to other similar software you have used, how would you rate this performance management system in terms of ease of use? 5: Much less complicated ~ 1: Much more complicated	34	3.65	29	3.52	5	4.40
Overall, how easy or difficult was the system to use? 5: Very easy ~ 1: Very difficult	34	3.82	29	3.72	5	4.40
Overall, how easy or difficult was the system to understand? 5: Very easy ~ 1: Very difficult	34	3.91	29	3.93	5	3.80
Overall, how professional or unprofessional did the system appear? 5: Very professional ~ 1: Very unprofessional	34	4.59	29	4.55	5	4.80
Overall, how efficient or inefficient was the system? 5: Very efficient ~ 1: Very inefficient	34	4.00	29	4.00	5	4.00
Overall, as you worked through the tasks, did the product become… 5: Much easier to use ~ 1: Much harder to use	34	4.18	29	4.10	5	4.60
Overall, how effective or ineffective do you think the performance management system will be as a career development and career planning tool? 5: Very effective ~ 1: Very ineffective	34	4.18	29	4.17	5	4.20

Table 14. Usability Posttest Survey Outcomes by Supervisor Status (continued)

Variable Description	Overall		Supervisor Status			
			Supervisors		Nonsupervisors	
	(n)	Avg. Score	(n)	Avg. Score	(n)	Avg. Score
I have a clear understanding of the performance management system. 5: Strongly agree ~ 1: Strongly disagree	34	3.59	29	3.66	5	3.20
The performance management system seems fair/accurate. 5: Strongly agree ~ 1: Strongly disagree	34	3.97	29	3.97	5	4.00
The performance management system allows performance reviews to be conducted in a timely manner. 5: Strongly agree ~ 1: Strongly disagree	34	3.74	29	3.69	5	4.00
I am satisfied with the test version of the performance management system. 5: Strongly agree ~ 1: Strongly disagree	34	3.76	29	3.76	5	3.80

[a] *The original agreement scale in both survey questionnaires had the opposite endpoints—1 indicated "strongly agree" and 5 "strongly disagree." This scale was reversed in the analysis for convenient presentation.*
[S] *significantly different from supervisors at the .05 level (t-test).*
[NS] *significantly different from nonsupervisors at the .05 level (t-test).*

Tables 15 and 16 compare the overall user ratings measured on pretest and posttest surveys from Usability Study I to Usability Study II. For pretest survey results, three items yielded statistically significant differences between test periods. Sailors participating in Usability Study II were significantly higher in their ratings of having a clear understanding of the present EVAL/FITREP system, being able to submit their own input at their last EVAL/FITREP, and the fairness/accuracy of their last advancement/promotion recommendation. For posttest survey results, four items yielded statistically significant differences. Sailors participating in Usability Study II were significantly higher in ratings of their comfort level performing usability tasks, the level of professionalism in the system's appearance, system efficiency, and the overall effectiveness of the performance management system as a career development and career planning tool. Furthermore, including all nonsignificant results, all posttest ratings were higher for Usability Study II.

Table 15. Usability Pretest Survey Outcomes by Study Period

	Overall		Usability I		Usability II	
Variable Description	**(n)**	**Avg. Score**	**(n)**	**Avg. Score**	**(n)**	**Avg. Score**
I have a clear understanding of the present EVAL/FITREP system. 5: Strongly agree ~ 1: Strongly disagree[a]	91	4.13	57	4.00[U2]	34	4.35[U1]
My last EVAL/FITREP was fair/accurate. 5: Strongly agree ~ 1: Strongly disagree	91	4.22	57	4.11	34	4.41
My last EVAL/FITREP was conducted in a timely manner. 5: Strongly agree ~ 1: Strongly disagree	91	4.05	57	4.00	34	4.15
I was able to submit my own input at my last EVAL/FITREP. 5: Strongly agree ~ 1: Strongly disagree	91	4.30	57	4.12[U2]	34	4.59[U1]
My last advancement/promotion recommendation was fair/accurate. 5: Strongly agree ~ 1: Strongly disagree	91	4.31	57	4.18[U2]	34	4.53[U1]
I am satisfied with the present Navy EVAL/FITREP system. 5: Strongly agree ~1: Strongly disagree	91	3.44	57	3.46	34	3.41
The most qualified and deserving Sailors score the highest on their EVALs/FITREPs. 5: Strongly agree ~ 1: Strongly disagree	91	3.27	57	3.25	34	3.32
How easy or difficult do you think it will be to use this test version of the performance management system? 5: Very easy ~ 1: Very difficult	87	2.91	53	2.96	34	2.82
How efficient or inefficient do you think the performance management system will be? 5: Very efficient ~ 1: Very inefficient	87	3.29	53	3.17	34	3.47

[a] *The original agreement scale in both survey questionnaires had the opposite endpoints—1 indicated "strongly agree" and 5 "strongly disagree". This scale was reversed in the analysis for convenient presentation.*

[U1] *significantly different from Usability Study I at the .05 level (t-test).*

[U2] *significantly different from Usability Study II at the .05 level (t-test).*

Table 16. Usability Posttest Survey Outcomes by Study Period

Variable Description	Study Period					
	Overall		Usability I		Usability II	
	(n)	Avg. Score	(n)	Avg. Score	(n)	Avg. Score
How comfortable or uncomfortable did you feel performing the tasks in the test? 5: Very comfortable ~ 1: Very uncomfortable	89	3.37	55	3.15^{U2}	34	3.74^{U2}
How certain or uncertain are you that you completed the tasks successfully? 5: Very certain ~ 1: Very uncertain	89	3.51	55	3.36	34	3.74
Compared to other similar software you have used, how would you rate this performance management system in terms of ease of use? 5: Much less complicated ~ 1: Much more complicated	88	3.42	54	3.28	34	3.65
Overall, how easy or difficult was the system to use? 5: Very easy ~ 1: Very difficult	89	3.60	55	3.45	34	3.82
Overall, how easy or difficult was the system to understand? 5: Very easy ~ 1: Very difficult	89	3.67	55	3.53	34	3.91
Overall, how professional or unprofessional did the system appear? 5: Very professional ~ 1: Very unprofessional	89	4.36	55	4.22^{U2}	34	4.59^{U2}
Overall, how efficient or inefficient was the system? 5: Very efficient ~ 1: Very inefficient	89	3.66	55	3.45^{U2}	34	4.00^{U2}
Overall, as you worked through the tasks, did the product become … 5: Much easier to use ~ 1: Much harder to use	89	4.08	55	4.02	34	4.18
Overall, how effective or ineffective do you think the performance management system will be as a career development and career planning tool? 5: Very effective ~ 1: Very ineffective	88	3.90	54	3.72^{U2}	34	4.18^{U2}

(continued)

Table 16. Usability Posttest Survey Outcomes by Study Period (continued)

Variable Description	Overall		Study Period			
			Usability I		Usability II	
	(n)	Avg. Score	(n)	Avg. Score	(n)	Avg. Score
I have a clear understanding of the performance management system. 5: Strongly agree ~ 1: Strongly disagree	89	3.34	55	3.18	34	3.59
The performance management system seems fair/accurate. 5: Strongly agree ~ 1: Strongly disagree	89	3.85	55	3.78	34	3.97
The performance management system allows performance reviews to be conducted in a timely manner. 5: Strongly agree ~ 1: Strongly disagree	89	3.67	55	3.64	34	3.74
I am satisfied with the test version of the performance management system. 5: Strongly agree ~ 1: Strongly disagree	89	3.58	54	3.46	34	3.76

[a] *The original agreement scale in both survey questionnaires had the opposite endpoints—1 indicated "strongly agree" and 5 "strongly disagree." This scale was reversed in the analysis for convenient presentation.*
[U1] *significantly different from Usability Study I at the .05 level (t-test).*
[U2] *significantly different from Usability Study II at the .05 level (t-test).*

7 Summary and Conclusions

7.1 Key Findings

Overall, the HPFD and ePerformance systems continued to function well in the Usability Study II. NSIPS functionality improved significantly from Usability Study I. Users were happier with the system tested in Usability Study II than with the system tested previously. Finally, the document workflow component succeeded—with the help of the QRG, users were able to find documents others had created for them and were able submit completed documents up the chain of command. Details on specific findings relating to the HPFD and ePerformance systems and overall user satisfaction are presented below.

7.1.1 HPFD System

The HPFD system testing provided the strongest data for verifying that changes made after Usability Study I truly resulted in system improvement. Some key findings for were

- Overall, tasks were completed more quickly during Usability Study II.
- The differences between supervisors and nonsupervisors and between locations were less pronounced in Usability Study II.
- Users at BUPERS were able to locate, create, and open documents most quickly. Completion times for these tasks may have been shortened by the introduction of an enhanced (more detailed) QRG at this site.
- An overall reduction in errors was observed from Usability Study I to Usability Study II. The most commonly observed "error" in Usability Study II was the participant's use of the QRG—not a usability problem *per se* but a demonstration of how important that solution to a Usability Study I issue was.
- Users aboard USS HOWARD (DDG 83) experienced shorter processing times due to the availability of a shipboard NSIPS server. The reduced system processing times seemed to be associated with fewer usability errors overall.
- Users had the most trouble locating and opening documents.

7.1.2 ePerformance System

In general, the ePerformance system testing had results similar to the HPFD system testing. Some key findings for the ePerformance system were

- Tasks were completed more quickly aboard USS HOWARD (DDG 83) due to the server situation described previously.
- Users experienced the same problems locating and opening documents as they experienced in HPFD system testing. Additionally, the ePerformance system required users to create some of their own documents. This task was equally challenging.

7.1.3 User Satisfaction

Overall, test users gave the systems positive ratings on the user surveys. Furthermore, supervisors and nonsupervisors were equally satisfied with the performance management/appraisal system. In general, Usability Study II participants were more satisfied with the system overall. Specifically, the posttest survey ratings in Usability Study II were significantly higher than in Usability Study I on the following dimensions:

- comfort level performing tasks in the test,
- level of professionalism the system conveyed,
- efficiency of the system, and
- overall effectiveness the system will have as a career development/planning tool.

7.1.4 Suggestions for Further Improvement

These key findings across HPFD and ePerformance systems, and overall system satisfaction, are useful for determining next steps for continued system improvement and maintenance, as well as the recommended approach for full-scale implementation. Based on these findings, it is clear that the NMCI connection speeds must be under constant surveillance by NPC to maintain top processing speeds and efficiency. This is critical to ensuring system functionality and user satisfaction. Additionally, users will require and should be provided with a hard copy QRG. The results show that users depend on this document to navigate the complex PeopleSoft file structure. Moreover, the QRG should be expanded to cover all HPFD and ePerformance system tasks.

7.2   Limitations of Research

The objective of usability testing is typically not to test for group differences to generalize findings to a population of users as a whole. Usability testing is one step in the HSI process that is usually followed by a larger pilot study with samples that more closely approximate the population of interest. Typically, usability testing employs small sample sizes using an iterative approach to calibrate a system or tool for pilot testing or implementation. These small sample sizes limit the generalizability and representativeness of the results.

Relative to other usability testing designs, this study used a reasonably large number of participants—34 Sailors across three iterations. However, for logistical reasons, these participants were heavily skewed to the supervisory population of the Navy. Few nonsupervisors were tested and, as a result, supervisory status dropped out as a significant variable affecting results when comparing Usability Study I to Usability Study II. As with any performance appraisal system, HPFD and ePerformance systems will be more heavily used by supervisors because these people will be responsible for many different types of performance management and appraisal tasks. It will be important for Navy performance appraisal stakeholders not to lose sight of the needs of nonsupervisors using the system, especially since they seemed especially vulnerable to error and frustration in Usability Study I.

7.3   Future Research

Results from this study raise several questions that could be addressed through future research. First, as recommended by several users in focus group interviews (Schwerin et al., 2006) a full-scale, command-wide pilot study of the HPFD and ePerformance systems, which responds to pre- and posttest user perceptions and recommendations for improvement, would provide a more generalizable evaluation of the two systems. This command-wide pilot study would also confirm that changes made to the system as a result of Usability Study II indeed made it a more usable system. For example, improved document workflow following Usability Study I resulted in fewer navigation problems, decreased time spent completing tasks, and increased satisfaction with the system.

Second, work should continue on the QRG, expanding it to cover all tasks within the HPFD and ePerformance systems. Specifically, the usability test results showed that users need instructions on how to enter a performance note. Other helpful features for tasks not tested

might include instructions on how to create HPFD documents and review ePerformance documents. Users also may benefit from a flow chart representing the path of documents through the system along with the in-person interaction between Sailors and their supervisors. Finally, a Frequently Asked Questions (FAQs) page might make it easier for users to find quick answers to questions about small tasks like how to print documents, check spelling, open and close all sections of a form. Revisions made to the QRG should be tested thoroughly with users. Any updates to the system made as a result of Usability Study II should be reflected in the QRG.

Finally, once HPFD and ePerformance are implemented Navy-wide, user perceptions of procedural fairness in performance appraisal should be collected on Navy personnel surveys (e.g., Navy-Wide Personnel Survey or Navy Quality of Life Survey) to evaluate the impact of the new HPFD and ePerformance system on organizational outcomes such as job satisfaction, commitment to the organization, and retention intent. These surveys provide key benchmarks by which the Navy can develop an understanding of the long-term impact of changes to the performance appraisal process. Existing workplace commitment/retention models (Michael & Olmsted, 2002) could be used to compare the impact of current performance appraisal system with HPFD/ePerformance. The positive findings of this study imply that Navy-wide implementation of the systems will likely improve perceptions of fairness, job satisfaction, and commitment to the organization. Collecting data on the full-scale transition to HPFD and ePerformance in Navy-wide personnel surveys and conducting analyses with existing models will provide further validation of these results.

8 References

American Psychological Association (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: American Psychological Association.

Blau, G. (1999). Testing the longitudinal impact of work variables and performance appraisal satisfaction on subsequent overall job satisfaction. *Human Relations, 52*(8), 1099-1113.

Borman, W.C., Hedge, J.W., Bruskiewicz, K.T., & Bourne, M.J. (2003). *Initial development work on an objective Navy advancement system linked to the Task Force EXCEL model*. Tampa, FL: Personnel Decisions Research Institute, Inc. (Institute Report #429)

Buff, W. (2004, August). *Human Engineering and Human Systems Integration*. Paper presented at the annual meeting of the American Psychological Association.

Daily, R.C., & Kirk, D.J. (1992). Distributive and procedural justice as antecedents of job dissatisfaction and intent to turnover. *Human Relations, 45*(3), 305-317.

Dean, E., Schwerin, M., Lee, S., Robbins, K., & Bourne, M. (2004). *Usability testing of the U.S. Navy Performance Management System: Technical report #2*. Unpublished Technical Report, Research Triangle Park, NC: RTI International.

Department of Defense (2003, May 12). *Department of Defense Instruction Number 5000.2: Operation of the Defense Acquisition System*. Washington, DC: Department of Defense. Retrieved January 4, 2006, from http://www.dtic.mil/whs/directives/corres/pdf2/i50002p.pdf

Diekmann, K. A., Barsness, Z. I., & Sondak, H. (2004). Uncertainty, fairness perceptions, and job satisfaction: a field study. *Social Justice Research, 17*(3), 237-255.

Dumas, J.S., & Redish, J.C. (1993). *A practical guide to usability testing*. Norwood, NJ: Ablex Publishing Corporation.

Ellickson, M.C., & Jogsdon, K. (2002). Determinants of job satisfaction of municipal government employees. *Public Personnel Management, 31*(3), 343-358.

Executive Review of Navy Training (August 2001). *Revolution in training: Final report*. Commissioned by the Chief of Naval Operations.

Giles, W.F., Findley, H.M., & Field, H.S. (1997). Procedural fairness in performance appraisal: Beyond the review session. *Journal of Business and Psychology, 11*(4), 493-506.

Hedge, J.W., Borman, W.C., Bruskiewicz, K.T., & Bourne, M.J. (2002). *The development of performance management systems and feedback tools for supervisory and nonsupervisory jobs in the U.S. Navy*. Minneapolis, MN: Personnel Decisions Research Institutes, Inc. (Institute Report #415)

Hedge, J.W., Bruskiewicz, K.T., Borman, W.C., & Bourne, M.J. (2004). *Preliminary identification of factors for Navy officer promotion linked to the Task Force Excel model*. Minneapolis, MN: Personnel Decisions Research Institutes, Inc. (Institute Report #453)

Jones, F.F. (1998). Pay procedures and voluntary turnover: Does procedural justice matter? *Psychological Reports, 83*(2), 475-482.

Lemons, M.A., & Jones, C.A. (2001). Procedural justice in promotion decisions: Using perceptions of fairness to build employee commitment. *Journal of Managerial Psychology, 16*(4), 268-280.

Levy, P.E., & Williams, J.R. (1998). The role of perceived system knowledge in predicting appraisal reactions, job satisfaction, and organizational commitment. *Journal of Organizational Behavior, 19*(1), 53-65.

Masterson, S.S. (2001). A trickle-down model of organizational justice: Relating employees' and customers' perceptions of and reactions to fairness. J*ournal of Applied Psychology, 86*, 594-604.

Mayer, R.C. & Davis, J.H. (1999). The effect of the performance appraisal system on trust for management: A field quasi-experiment. *Journal of Applied Psychology, 84*(1), 123-136.

Michael, P.G., & Olmsted, M.G. (2002). *Affective organizational commitment, retention intent, and retention behavior: Investigation of a path model*. Paper presented at the Annual Convention of the Society for Industrial and Organizational Psychology, Ontario, Canada.

Nielsen, J. (1993). *Usability engineering*. Boston: AP Professional.

Norman, K.L, Friedman, Z., Norman, K., & Stevenson, R. (2000). *Navigational issues in the design of on-line self-administered questionnaires*. College Park: University of Maryland. Human-Computer Interaction Laboratory. Retrieved January 31, 2006, from http://hcil.cs.umd.edu/trs/2000-10/2000-10.pdf

Olmsted, M. & Underhill, C. (2003). *2000 Navy-wide Personnel Survey (NPS): Report of survey findings* (TN-03-11). Millington, TN: Navy Personnel Research, Studies, and Technology.

PeopleSoft, Inc. (2004). *PeopleSoft 8.8*. Pleasanton, CA: PeopleSoft, Inc.

Schwerin, M.J., Dean, E. & Aspinwall, K.R. (2006). *Qualitative data from Human Performance Feedback and Development (HPFD) and ePerformance system users: Technical Report #2*. Unpublished Technical Report, Research Triangle Park, NC: RTI International.

Schwerin, M., Dean, E., Robbins, K., & Bourne, M. (2004). *Usability testing of the U.S. Navy Performance Management System: Technical report #1*. Unpublished Technical Report, Research Triangle Park, NC: RTI International.

Simons, T., & McLean Parks, J. (2000, August). *The sequential impact of behavioral integrity on trust, commitment, discretionary service behavior, customer satisfaction, and profitability*. Paper presented at the National Conference of the Academy of Management, Toronto, Ontario, Canada.

Westerman, S.J. (1997). Individual differences in the use of command line and menu computer interfaces. *International Journal of Human-Computer Interaction*, 9(2), 183-198.

Wiedenbeck, S. (1999). The use of icons and labels in an end user application program: an empirical study of learning and retention. *Behavior & Information Technology*, 18(2), 68-82.

Appendix A:

Usability Error Codes used for Usability Study I and Usability Study II

Code	Error	Explanation
C	Confused	Participant looks confused—furrowing of eyebrows, quizzical expressions
D	Doesn't Follow Screen Instructions	Participant does not follow on-screen instructions needed to complete task.
E	Incorrect Click on Wrong Button	Participant clicks on the wrong button while attempting to complete a PeopleSoft task.
F	Failed Task	Participant is unable to complete task.
G	General Button Error	Participant clicks on button and it does not work or gives participant an error message.
H	Asks for Help	Participant looks for reassurance or specifically asks for help.
I	Refers to Quick Reference Guide (QRG)	Participant looks at QRG.
K	"I don't know how to do that."	Participant says they don't know how to complete the task, or can't do it, and noticeably gives up on the task at hand.
M	System Server Problem	NSIPS server shuts down and does not function.
N	Navigational Error	Participant clicks on wrong folders or follows wrong path while trying to create, open, or find a document.
O	Attempt to Search Outside Document	Participant searches for a button or a function outside of the HPFD or ePerfromance document (such as in the Internet Explorer menu bar or Windows Help Menu.)
Q	Boredom	Participant looks bored
R	Frustration	Participant looks frustrated
T	Time Out	Participant receives "connection has expired" or "page not available" screen.
Y	User Retries Action	Participant retries action while waiting for something to happen.