Contract Number: 4TNP17041031
RTI Project Number: 09166.002

Usability Testing of the U.S. Navy Performance Management System: Technical Report #2

Submitted to:

Navy Personnel Command
ATTN: CDR Mark J. Bourne, MSC, USN
5720 Integrity Drive
Millington TN 38055

Authors:
Elizabeth Dean, M.A.
Michael J. Schwerin, Ph.D.
Sunghee Lee, Ph.D.
Kimberly M. Robbins, M.A.
RTI International
3040 Cornwallis Road
Research Triangle Park, NC 27709-2194

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

CDR Mark J. Bourne MSC, USN
5720 Integrity Drive
Millington TN 38055

October 13, 2004

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
13 OCT 2004		**00-00-2004 to 00-00-2004**

4. TITLE AND SUBTITLE **Usability Testing of the U.S. Navy Performance Management System: Technical Report #2**	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **RTI International,3040 Cornwallis Road,Research Triangle Park,NC,27709-2194**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR’S ACRONYM(S) 11. SPONSOR/MONITOR’S REPORT NUMBER(S)

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

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**43**	

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

Table of Contents

List of Tables

Executive Summary

Through work with the U.S. Navy's Task Force for Excellence through Commitment to Education and Learning (EXCEL), the Navy has begun the process of aligning Fleet personnel requirements with training, manpower and personnel processes. A key component of Task Force EXCEL was the development of a web-based, performance management system that focuses on workplace behaviors rather than the Navy's current trait based system.

One of challenges for the Performance Vector was to develop a performance management and performance appraisal system that is aligned with the changing workplace performance needs of the U.S. Navy. Since 1996, the Navy has operated with a trait-based, performance system, where supervisors have rated personnel on traits such as leadership, teamwork, equal opportunity, and military bearing/character (BUPERSINST 1600.10, 1995. The first step in meeting this challenge was the development of the Human Performance Feedback and Development (HPFD) model – a behaviorally based performance management system with dimensions for supervisory and non-supervisory personnel that reflect those qualities that Navy leaders endorse as essential for maintaining a high-quality Navy workforce.

The objective of this study was to assess a pilot version of the Web-based, behaviorally based, performance management (HPFD) and appraisal system (ePerformance). Specifically, the objectives were to capture quantitative and objective data as well as qualitative and subjective data from participants to identify potential sources of error and user burden. To capture these data, the Performance Vector Research Team (PVRT) and RTI International conducted usability testing and user surveys. This study assessed problems and errors using the Web-based system and the users' perceptions about the proposed new system.

Data collection took place in three iterations at three different locations: Naval Air Station (NAS) Brunswick, USS KITTY HAWK (CV63), and Naval Base Kitsap – Bangor. Usability scenarios conducted at each site evaluated the effectiveness of screen layouts, performance item structures, and on-screen features for the Navy's HPFD and ePerformance systems. Two paper-and-pencil self-administered surveys—pre-test and post-test surveys—were administered to obtain Navy personnel's subjective impressions of the HPFD and ePerformance systems.

The results show that users experienced functional problems using the Navy's Standard Integrated Personnel System (NSIPS), specifically system timing out and long page loading times. Overall, results from testing indicate that the HPFD and ePerformance systems worked well. The usability survey results suggest no major systematic differences of perception between supervisory and non-supervisory users although non-supervisors did experience a slightly higher rate of usability errors.

Results also indicate that system errors had a significant negative effect on users' ability to access documents, use the system, and their satisfaction with the HPFD and ePerformance modules. All efforts should be made to increase the speed of the system and to decrease the occurrence of the system timing-out. Improvements to NSIPS that enable consistent and reliable access to the HPFD and ePerformance documents are likely to greatly enhance the system usability and user satisfaction. Future rounds of testing the Web-based, performance management system should focus on vertical document workflow, usability among non-supervisory personnel, and the effect of changes made to the HPFD and ePerformance appraisal system identified from this study.

1 Introduction

The Chief of Naval Operations (CNO) chartered the Executive Review of Navy Training (Clark, 2001), which subsequently led to the formation of a Task Force for Excellence through Commitment to Education and Learning (EXCEL). Task Force EXCEL's goal was to identify new ways for the U.S. Navy to train, grow, place, and utilize personnel who maximize the Navy's ability to accomplish its military mission while developing a more productive yet satisfying workplace.

Task Force EXCEL consists of five components or "*vectors*" that are essential to how personnel meet their missions and manage the Navy workforce. These five vectors include Professional Development, Personal Development, Professional Military Education and Leadership, Certifications and Qualifications, and Performance. The primary tasking of the Performance Vector includes an examination of the Navy performance appraisal and management system.

One challenge for the Performance Vector was the need for a performance appraisal and management system that is aligned with the changing workplace performance needs of the U.S. Navy. Since 1996, the Navy has operated with a trait-based performance appraisal system, in which supervisors have rated personnel on traits such as leadership, teamwork, equal opportunity, and military bearing/character (BUPERSINST 1600.10, 1995). One recommendation from initial Task Force EXCEL meetings was a behaviorally based performance appraisal system. In addition, after examining military and civilian best practices in performance appraisal and management, and learning of the CNO's desire for an electronically based performance management/appraisal system, the Performance Vector recommended the development of a behaviorally based performance appraisal system.

The Commander, Navy Personnel Command (CNPC) is confronted with having to develop performance appraisal systems that are fully operational and integrated with the performance evaluation and promotion selection cycle. As the new Human Performance Feedback and Development (HPFD) performance management and appraisal system is implemented, the final performance appraisal forms as formatted and presented in the PeopleSoft 8.8 (2004) application require usability testing with supervisory and non-supervisory Navy personnel to identify usability concerns and improve the functionality of the electronically based performance management and appraisal system. Usability testing is a vital step in the development of any new Web-based tool. In theory, the automated tool should reduce the burden on users. In practice, however, such tools can be more difficult to figure out than their paper counterparts. Usability testing can assess the time it takes to complete a form, the amount of self-editing required, and the navigational problems users face. It can also assess users' emotive reactions to instruments. Identifying sources of burden and reducing the causes of user stress result in a more efficient Web-based system.

The objectives for this study were to capture quantitative and objective data as well as qualitative and subjective data from participants to identify potential sources of error and user burden. To capture these data, usability testing was combined with user pretest and post-test surveys to assess problems and errors using the Web-based system as well as the users' perceptions about the proposed new system in relation to the existing performance appraisal process.

2 Literature Review

2.1 Performance Appraisal Systems

The impact of performance appraisal systems on job satisfaction, organizational commitment, and retention has been a topic of research among civilian researchers for several years. Daily and Kirk (1992) examined perceptions of workplace/procedural fairness and demonstrated a strong relationship between workplace fairness (including variables associated with procedural justice and satisfaction with the performance appraisal process) and voluntary turnover intent. Levy and Williams (1998), after controlling for actual performance ratings, demonstrated that performance appraisal satisfaction and perceived system knowledge have a strong, significant relationship with job satisfaction and organizational commitment. When examining the effect of work factors on retention plans, Jones (1998) found, after controlling for the effects of demographic characteristics and distributive justice, the perceived fairness of procedures for pay determination, performance appraisals, and appeals were related to voluntary turnover. More recent studies (Blau,1999; Ellickson & Jogsdon, 2002) demonstrated a strong statistical relationship between work life factors (including measures of satisfaction with the performance appraisal process) and job satisfaction among those in the civilian workforce.

Descriptive analyses of Sailors' satisfaction with their current performance appraisal system indicate 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). While over half of enlisted personnel (58%) and over three-fourths of officers (77%) believed their current performance appraisal system was "fair and accurate," only 29% of enlisted personnel and 49% of officers believed that “the most qualified and deserving Sailors rank high on their EVALs/FITREP.” [1] In addition, while a majority of enlisted personnel and officers reported that they understand the advancement and promotion system (76% of enlisted personnel; 83% of officers), only 31% of enlisted personnel and 50% of officers were “satisfied with the present Navy advancement and promotion system.” Merely 20% of enlisted personnel and 41% of officers believed that “the most qualified and deserving Sailors get advanced or promoted.” These results would seem to suggest only a small endorsement of the Navy’s current performance management and appraisal systems.

The first study in this system development program (Hedge, Borman, Bruskiewicz, & Bourne, 2002) resulted in the development of the Human Performance Feedback and Development (HPFD) model—a behaviorally based job performance management system with dimensions for supervisory and non-supervisory personnel that reflect the qualities that Navy leaders endorse as essential for maintaining a high-quality Navy workforce. Subsequent research (Borman, Hedge, Bruskiewicz, & Bourne 2003; Hedge, Bruskiewicz, Borman, & Bourne, 2004) has identified the relative strength of performance dimensions at career stages for Navy enlisted personnel (recruit-apprentice, apprentice-journeyman, and journeyman-master) and officers (junior, mid-grade, and senior). This work culminated in the development of a Web-based HPFD system and a Web-based ePerformance system using a commercially available performance management software system, PeopleSoft Version 8.8 (PeopleSoft, 2004). These systems are housed within the Navy Standard Integrated Personnel Systems (NSIPS)—a secure Web-based environment that holds several Navy personnel systems.

[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.

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. While 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. The Undersecretary of Defense for Acquisition, Technology, and Logistics (USD AT&L) recently issued DoD Instruction 5000.2 (DODINST, 2003) that 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" (Enclosure 7, paragraph E7.1.1, p 43). It is clear that it is 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.

This study examines the usability of the Navy's pilot HPFD and ePerformance performance appraisal systems. 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. ***Section 2.2*** provides a review of the current best practices for usability testing.

2.2 Usability Testing

A succinct definition of usability testing is found in Dumas and Redish's (1993) handbook, *A Practical Guide to Usability Testing*. The authors note that since the primary goal of usability testing is to improve the usability of the product, specific goals and concerns need to be articulated when planning each test. For example, for the usability testing of the Navy's Web-based performance management tool, a specific goal was to assess the different usability needs for supervisors and non-supervisors and for shipboard and non-shipboard Sailors. In a usability test, the following four key factors must be present:

- The participants represent real users.
- The participants do real tasks.
- The usability researcher observes and records what participants do and say.
- The usability researcher analyzes the data, diagnoses the problems, and recommends changes to fix the problems (Dumas & Redish, 1993).

Nielsen (1993, p. 165) describes usability testing as "the most fundamental usability method" and "irreplaceable," because it's the only mechanism that allows the researcher to obtain direct, detailed information on users' experience with the product or tool being tested.

Usability researchers agree that multiple methodologies can effectively assess the user experience. In fact, most usability test plans include several types of data collection. Methods include baseline tests of existing products to assess pre-existing problems, surveys of user needs, user focus groups, participatory design experiences, heuristic evaluations, task analysis, and paper prototyping. The two most consistently emphasized assessment practices are an iterative design and consideration of user context.

2.2.1 Iterative Design

In a survey of usability researchers, Nielsen (1993) identified the most effective six methods for usability improvement. Iterative design (tied with task analysis) was the number one consideration. There are several reasons why iterative design of usability tests is so important. Changes to a system as a result of usability testing sometimes do not solve a problem. In fact, new solutions may create new problems. Furthermore, new solutions may reveal additional problems that were previously hidden or outbalanced by the original problem identified. Nielsen's research analyzing the effectiveness of iterative testing found a median improvement in system usability, defined by the usability metrics employed for the particular test plan, of 38% per iteration. While five out of 12 iterations in Nielsen's analysis showed that one dimension of usability had gotten worse, significant improvements in usability continued to be made in later iterations.

In the early days of usability testing (the 1970s and 1980s), the norm was one large-scale test of 30 users, conducted very late in the design process when most of the design features were stabilized and thus averse to change. The problem with this approach was that it found pervasive system problems, but at a stage in the development cycle where it was too late to fix them. In addition, 30 users were not needed to identify such large and pervasive problems. The solution adopted was to test earlier prototypes of systems, even using paper prototypes when necessary, with multiple iterations of five to 10 users. This approach allows early identification of large-scale systemic problems. Since 1990, iterative testing with small samples has been the preferred approach (Dolan & Dumas, 1999).

2.2.2 Context Awareness

Valid usability measurement cannot take place outside the user's context, and usable systems require incorporating this context into the development cycle. When considering tools such as guidelines and checklists for user-centered design, Bevan and Macleod (1994) warn against dependence on checklists, because guidelines for usable system features need extensive detail to be useful, but if checklists are detailed enough, they are likely to be too specific to apply in multiple real-world contexts. For example, a highly interactive Web-based performance management evaluation form that requires frequent communication with a server to complete may be desirable in an office setting because it will allow the user's data to be saved through many interruptions. Conversely, this approach may not be desirable on board a deployed Navy ship, since the satellite Internet connection may be unavailable or regularly interrupted. The solution is to conduct scenario-based assessments that reflect the environments of real users. A true-to-life environment can be replicated in a lab setting, but the most realistic approach is to conduct on-site usability testing in the field. Bevan and Macleod add a fifth factor to Dumas and Redish's list above: The participant's real-life context is represented in the usability test.

This evaluation of the Navy's Web-based performance management system incorporated the two key design features of iterative testing and context awareness. Usability tests were conducted at three very different Navy installations with time between iterations to make changes to the system.

3 Study Objectives

The objectives for this study were to capture quantitative and objective data as well as qualitative and subjective data from participants to identify potential sources of error and user burden. Specifically, the objectives of this study were to conduct:

- Usability tests of the HPFD system with non-supervisory Navy personnel 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 of the HPFD and ePerformance systems with supervisory Navy personnel collecting data on the type and frequency of user errors, user reactions to the systems, and self-reported user satisfaction with the systems.
- User pretest and post-test surveys of non-supervisory and supervisory personnel who completed HPFD and ePerformance usability tests to identify expectations and overall satisfaction with the 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 the informed consent form and to return the form to their respective task leaders. No adverse events occurred during the course of this study.

4 Participants

The project manager identified a local, on-site liaison to assist in participant recruiting, scheduling, and study logistics. Instructions sent to the on-site liaison described the criteria for selecting potential participants—supervisory and non-supervisory personnel assigned to operational and shore commands or units, ranging in paygrade from E-2 through O-6.

4.1 Iteration 1: Naval Air Station (NAS) Brunswick

Iteration 1 took place at NAS Brunswick in Brunswick, Maine, from June 21, 2004, through June 25, 2004. A total of 21 active duty Navy personnel took part in data collection. All 21 personnel participated in the usability testing and completed the pre- and post-test usability surveys. Of the 21 personnel, 14 were supervisors, and seven were in non-supervisory positions. Ten participants were NAS Brunswick personnel, 10 were squadron personnel, and one participant was from a ship pre-commissioning unit. Only one of these personnel could not participate in the subsequent focus group interview.

4.2 Iteration 2: USS KITTY HAWK (CV63)

Iteration 2 took place aboard the USS KITTY HAWK (CV63) in Yokosuka, Japan, from July 12, 2004, through July 16, 2004. A total of 20 active duty Navy personnel were scheduled to part in data collection. Seventeen personnel participated in the usability testing and completed the pre- and post-test usability surveys. One participant could not be tested because the online system was unavailable, and another participant could not be tested because the ship's T1 line was disconnected to switch over to a satellite Internet connection. A third participant could not make the usability session because shipboard duties created a scheduling conflict. Of the 20 personnel, 14 were supervisors, and six were in non-supervisory positions. Seven were officers, and 13 were enlisted. Participants were from the following departments: six from the Air

Department, five from Air Intermediate Maintenance Department (AIMD), four from the Operations Department, two from the Combat Systems Department, one from Executive Officer Administration, one from Supply, and one from Weapons.

4.3 Iteration 3: Naval Base Kitsap – Bangor

Iteration 3 took place at the Trident Training Facility at Naval Base Kitsap in Bangor, Washington, from August 9, 2004, through August 13, 2004. A total of 20 active duty Navy personnel were scheduled to take part in data collection. Nineteen personnel participated in the usability testing and completed the pre- and post-test usability surveys. Of the 19 personnel, 10 were supervisors, and nine were in non-supervisory positions. All Navy personnel were enlisted Sailors. Participants were from the following commands: six from the USS ALABAMA (SSBN 731), six from the USS ALASKA (SSBN 732), one from the USS NEVADA (SSBN 733), four from the USS KENTUCKY (SSBN737), one from Commander Submarine Squadron Nineteen (CSS-19), and one from Commander Submarine Squadron Seventeen (CSS-17).

5 Instruments and Procedures

5.1 Usability Scenarios

Usability scenarios were developed to evaluate the effectiveness of screen layouts, performance appraisal item structures, and on-screen features for the Navy's HPFD and ePerformance systems. Specifically, usability testing protocol and scenarios targeted the following potential problems:

- *Unclear navigational instructions.* Are respondents able to tell where on the screen to start reading and where to supply the required information?
- *Confusing help text.* Is help text consistently displayed within the documents, and does the help text answer the users' most common questions?
- *Meaningless or confusing error messages.* Are error messages appropriately displayed when problems occur? Do respondents find the error messages informative and helpful rather than alarming or confusing?
- *Problems of accessing/responding via the Web.* What is the most efficient Web tool design for the least capable information technology (IT) platform and least advanced hardware and software?

Test scenarios were also developed to simulate actual tasks that Navy non-supervisors and supervisors are likely to encounter.

In an effort to test the HPFD and ePerformance systems in the field, this research study utilized a portable usability lab—a coordinated system of digital audio and video data capture equipment. The portable usability lab features professional grade video monitoring and recording capabilities, including two high-resolution video cameras with silent remote control pan, tilt, zoom, and focus.

Following the best practices in usability testing described above, an iterative approach with three separate rounds of usability testing was used. In order 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 warfare communities (i.e., surface, submarine, and aviation communities) in an Atlantic Fleet (i.e., NAS Brunswick), Pacific Fleet (Naval Base Kitsap – Bangor), and overseas (USS KITTY HAWK [CV63]—Yokosuka, Japan) locations.

5.2 Usability Survey

Two paper-and-pencil self-administered surveys—pre-test and post-test 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/management tool.

The pre-test 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 post-test survey asked participants to report their perceptions about completing the tasks in the usability 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.

6 Results

When analyzing the survey data and the usability test data, we used four independent variables: supervisor status (supervisor/non-supervisor), test site location (Naval Base Kitsap – Bangor, USS KITTY HAWK [CV63], or NAS Brunswick), current paygrade, and years served in the Navy. The first two variables were included because they are key variables of interest for the study. The last two were selected from a preliminary analysis that examined the correlation among paygrade, years in the Navy, education, and age. The correlation matrix indicated a strong relationship between years in the Navy and age at 0.76 and between paygrade and education at 0.72, both with *p-value* less than 0.01. Paygrade and years in the Navy may be considered as proxy measures of education and age, respectively. Other demographic variables, such as gender, race, and ethnicity, were considered initially but because little variation was observed in those variables they were excluded from further analyses.

Two analytic techniques were mainly used throughout the study, analysis of variance (ANVOA) and Chi-square tests of significance. ANOVA was used for continuous variables, such as task time and error frequency, and ordinal variables, such as those using the five-point agreement scale. For categorical variables, we used a contingency table. The differences in continuous and ordinal variables within the demographic variables introduced above were investigated with Bonferroni *t*-test that accounts for multiple comparisons. Since the Bonferroni *t*-test is a more stringent test of significance for between groups mean score comparisons, Tukey's *t*-test was used to determine if a less stringent test would affect the results. Tukey's *t*-test for group comparisons produced the same results. For categorical variables, we used Chi-square tests of significance for group differences. Although these tests require random, normally distributed samples, cautiously applying this statistical test 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 manipulate experimental usability stimuli to compare the effect of 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 but rather usability was examined between user groups (i.e., supervisors and non-supervisors 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 generalized to either the general population or to the Navy population. Generalization may be possible only through large-scale studies employing probability samples of the study target population. Second, since this study did not have experimental and control conditions, the associations between the independent and dependent variables should be viewed as correlational rather than causal.

6.1 Task Durations

An examination of the average completion times required for each task provides initial information on the relative demands placed on the users between supervisors/non-supervisors and among users at each of the three locations. Longer average completion times may be an indicator of increased burden. ***Table 1*** displays the estimates of the average completion time (presented in seconds) for each usability task as well as the results of significance tests of the differences among groups.

Table 1. Estimate of Average Time to Complete Usability Testing Task by Task[1, 2]

Task Description	(n)	Overall	Supervisor Status		Location		
			Supervisor (S)	Nonsup. (NS)	Naval Base Kitsap (K)	USS KITTY HAWK (Y)	NAS Brunswick (B)
Task1: Complete the CBT Tutorial.	26	1633.6	1822.5	1444.6	1480.1	1797.1	1641.7
Task2: Log in to NSIPS.	38	323.8	317.0	332.4	247.5	539.5 ^{B}	135.6 ^{Y}
Task3: Open the HPFD document.	45	212.0	209.8	216.1	209.8	213.6	213.1
Task4: Complete the HPFD document.	51	716.3	732.3	695.1	742.9	633.4	747.9
Task5: Check spelling.	38	88.1	88.3	87.8	78.0	103.8	85.6
Task6: Find the "Target Behaviors" description.	36	50.1	45.4	56.6	70.4 ^{B}	71.4 ^{B}	31.3 ^{Y}
Task7: Change ratings and cut and paste comments.	35	89.3	75.0 ^{NS}	120.6 ^{S}	102.6	88.8	80.4
Task8: Collapse all sections of the document.	41	32.1	30.5	34.5	45.4	30.5	23.1
Task9: Submit the HPFD document.	38	63.7	73.7	44.5	58.9	85.0	55.4
Task10: Enter a performance note.	39	143.5	135.1	154.4	168.7	175.0	108.2

[1] Time was measured by second.

[2] Tasks 11 through 19 were completed only by study subjects with supervisor status. Therefore, these tasks were excluded from the analysis.

Note: Superscripts ^{S}, ^{NS}, ^{K}, ^{Y}, ^{B} indicate significantly different estimates at the 0.05 level from *t*-test. Bonferroni *t*-test was used to account for multiple comparisons for location variable.

Only three tasks yielded statistically significant differences across groups: logging into NSIPS (Task 2), finding the “target behaviors” description (Task 6), and changing ratings and cutting and pasting comments (Task 7).

- Sailors on the USS KITTY HAWK (CV63) had the longest durations for logging into NSIPS, an average of 539.5 seconds compared to 247.5 seconds at Naval Base Kitsap – Bangor and 135.6 seconds at NAS Brunswick. The extended log on durations for USS KITTY HAWK (CV63) appears to be the result of multiple server problems at the site. The server was frequently down and users could not log in. Additionally, the ship shifted from wire to satellite communications on Day 4 of data collection, as the ship prepared to go to sea the following week.
- There is no apparent reason for the differences in time for finding the “Target Behaviors” button shown by Sailors at NAS Brunswick (31.3 seconds) compared to the USS KITTY HAWK (CV63) (70.4 seconds) and Naval Base Kitsap – Bangor (71.4 seconds) participants.
- The difference in time spent collapsing sections of the document is not apparent among locations but is visible between supervisors and non-supervisors. Supervisors took an average of 75 seconds to change ratings and cut and paste, whereas non-supervisors took an average of 120.6 seconds.

Overall, differences between supervisors and non-supervisors and among the three sites are minimal.

6.2 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. ***Table 2*** presents the estimates of the *total error frequency*—that is, the total number of errors across all types of error. Total error frequency varies from task to task because the amount of time it takes to complete each task as well as the complexity of each task varies. Most notably in ***Table 2***, there are no statistically significant differences between supervisors and non-supervisors. Differences among locations are statistically significant across locations for completing the computer-based training (CBT) tutorial (Task 1) and completing the HPFD document (Task 4). Errors on the HPFD CBT (Task 1) were significantly higher at Naval Base Kitsap - Bangor and on the USS KITTY HAWK (CV63) than at NAS Brunswick. Errors completing the HPFD document (Task 4) were significantly higher among Sailors at Naval Base Kitsap – Bangor than Sailors aboard the USS KITTY HAWK (CV63) and NAS Brunswick. As with the durations, results indicate no major pattern across locations or across supervisor status.

Table 3 illustrates the *rate* of error occurrence for each task. The rate of error occurrence is the percentage of cases in which errors occurred in each task. 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 that could be skewed by particularly problematic cases. Errors occurred 100% of the time during the CBT tutorial (Task 1), but only about 23% of the time when users were required to change ratings and cut and paste comments (Task 7). Analyses indicate statistically significant differences in the rate of errors for only two tasks for supervisors and non-supervisors: changing ratings and cutting and pasting comments (Task 7) and entering a performance note (Task 10). Both tasks had higher

Table 2. Estimate of Error Frequency by Task[1]

Task Description	(n)	Overall	Supervisor Status		Location		
			Supervisor	Nonsup.	Naval Base Kitsap (K)	USS KITTY HAWK (Y)	NAS Brunswick (B)
Task 1: Complete the CBT Tutorial.	26	21.38	24.00	18.77	26.20 [B]	26.50 [B]	12.00 [K,Y]
Task 2: Log in to NSIPS.	38	3.63	3.90	3.29	4.29	3.64	2.50
Task 3: Open the HPFD document.	45	4.98	4.55	5.75	4.81	2.83	6.65
Task 4: Complete the HPFD document.	51	2.33	1.76	3.09	4.50 [Y,B]	1.36 [K]	1.33 [K]
Task 5: Check spelling.	38	1.89	1.48	2.41	2.67	0.67	2.10
Task 6: Find the "Target Behaviors" description.	36	1.11	0.95	1.33	1.71	1.20	0.84
Task 7: Change ratings and cut and paste comments.	35	0.69	0.46	1.18	0.91	0.38	0.69
Task 8: Collapse all sections of the document.	41	1.22	1.38	1.00	1.86	0.75	0.95
Task 9: Submit the HPFD document.	38	1.53	1.80	1.00	1.50	1.00	1.87
Task 10: Enter a performance note.	39	1.95	1.77	2.18	2.80	1.71	1.29

[1] Tasks 11 through 19 were completed only by study subjects with supervisor status. Therefore, these tasks were excluded from the analysis.

Note: Superscripts [K], [Y], [B] indicate significantly different estimates at the 0.05 level. Bonferroni *t*-test was used to account for multiple comparisons for location variable.

Table 3. Estimate of Percentage of Error Occurrence by Task[1]

			Supervisor Status		Location		
Task Description	(n)	Overall	Supervisor	Nonsup.	Naval Base Kitsap (K)	USS KITTY HAWK (Y)	NAS Brunswick (B)
Task 1: Complete the CBT Tutorial.	26	100.0	100.0	100.0	100.0	100.0	100.0
Task 2: Log in to NSIPS.	38	92.1	90.5	94.1	100.0	84.6	87.5
Task 3: Open the HPFD document.	45	93.3	93.1	93.8	100.0	83.3	94.1
Task 4: Complete the HPFD document.	51	64.7	58.6	72.7	93.8*	50.0*	52.4*
Task 5: Check spelling.	38	79.0	71.4	88.2	100.0*	44.4*	85.0*
Task 6: Find the "Target Behaviors" description.	36	69.4	57.1	86.7	100.0	60.0	63.2
Task 7: Change ratings and cut and paste comments.	35	22.9	12.5*	45.5*	45.5	12.5	12.5
Task 8: Collapse all sections of the document.	41	58.5	54.2	64.7	85.7*	12.5*	57.9*
Task 9: Submit the HPFD document.	38	73.7	72.0	76.9	100.0*	44.4*	66.7*
Task 10: Enter a performance note.	39	61.5	45.5*	82.4*	93.3*	71.4*	29.4*

[1] Tasks 11 through 19 were completed only by study subjects with supervisor status. Therefore, these tasks were excluded from the analysis.

Note: Superscript * indicates a significant association between the percentage of error occurrence and the independent variable at the 0.05 level from chi-square test.

error rates for non-supervisors than for supervisors. Conversely, five of the tasks had statistically significant differences among locations in rate of errors. Sailors at Naval Base Kitsap – Bangor performed with a higher rate of error on the following five tasks:

- Complete the HPFD document (Task 4);
- Check spelling (Task 5);
- Collapse all sections of the document (Task 8);
- Submit the HPFD document (Task 9); and
- Enter a performance note (Task 10).

The statistically significant differences between participants aboard the USS KITTY HAWK (CV63) and those at NAS Brunswick were less consistent. Performance among participants on the USS KITTY HAWK (CV63) showed a higher error rate than performance among Sailors at NAS Brunswick for entering a performance note (Task 10), but on all other tasks, NAS Brunswick had the higher error rate.

Table 4 shows the most frequently occurring error category for each task. Timing out was the most frequent problem overall, appearing as the most frequent type of error in nine tasks. One reason for system timing out for these tasks was that HPFD completion tasks required data be sent from the local machine to the NSIPS server. Resetting the time-out duration and including a time-out indicator could reduce the number of timing out errors. The most common error with the remaining tasks was the navigational error. It is likely that navigational errors occurs more frequently when opening the HPFD document (Task 3), opening the ePerformance Appraisal document (Task 13), and entering a performance note (Task 10). All three of these tasks require the user to find a specific document within the PeopleSoft menu structure which is not intuitive to novice PeopleSoft users.

Table 4. Most Frequently Occurred Error by Task and Estimate of Its Average Frequency

Task Description	(n)	Average Frequency	Error Description
HPFD Tasks			
Task 1: Complete the CBT Tutorial.	26	17.19	Doesn't follow screen instruction
Task 2: Log in to NSIPS.	38	2.16	Can't set new passwords
Task 3: Open the HPFD document.	45	2.11	Navigation error
Task 4: Complete the HPFD document.	51	0.65	Refer to info sheet
Task 5: Check spelling.	38	0.87	Time out
Task 6: Find the "Target Behaviors" description.	36	0.64	Time out
Task 7: Change ratings and cut and paste comments.	35	0.17	PeopleSoft button error
Task 8: Collapse all sections of the document.	41	0.63	Time out
Task 9: Submit the HPFD document.	38	0.39	Time out
Task 10: Enter a performance note.	39	0.64	Navigation error
ePerformance Tasks			
Task 11: Log out of PeopleSoft.	5	-	Nothing particular
Task 12: Log into NSIPS using ePerformance test account.	7	0.43	Can't set new passwords
Task 13: Open the Annual Performance Appraisal 1 document.	7	3.57	Navigation error
Task 14: Complete the Annual Performance Appraisal 1 document.	21	-	Nothing particular
Task 15: Check the ratings descriptions for one dimension.	12	0.25	Time out
Task 16: Check spelling.	11	0.36	Time out
Task 17: Check language.	18	0.39	Time out
Task 18: Calculate ratings.	11	0.37	Time out
Task 19: Submit the Annual Performance Appraisal 1 document.	7	0.71	Time out

Note. EPerformance Tasks (Tasks 11 through 19) were completed only by study subjects with supervisor status.

6.3 User Ratings

User ratings measured on pre-test and post-test surveys were compared between supervisors/non-supervisors and among test sites using ANOVA. Because of power limitations due to low sample size typical of usability studies, only supervisor/non-supervisor analyses appeared to have enough power to explain group differences on the dependent variables. As a result, analyses on user ratings will focus on comparisons between supervisory and non-supervisory user ratings. ***Table 5*** presents variable mean scores and the difference of means between supervisors and non-supervisors.

Items on the pre-test survey presented a series of questions about the user's satisfaction with the current EVAL/FITREP system and their expectations for using the Web-based system. Results from analyses of the pre-test survey indicate significant differences on five aspects of Sailors' perceptions of the EVAL/FITREP system. Supervisors are significantly higher in their ratings of having a clear understanding of the FITREP/EVAL system, fairness/accuracy, timeliness, Sailors submitting their own input, and perceptions of fairness in advancement/promotion. As ***Table 5*** shows, items reflecting user expectations of using the new system do not show statistically significant differences between supervisors and non-supervisors with a mean score indicating users believe the Web-based system will be "neither easy nor difficult."

Items on the post-test survey asked users how they felt about the test version of the Web-based performance management system. Supervisor ratings were significantly higher for three aspects of using the Web-based HPFD/ePerformance appraisal system – comfort performing the HPFD/ePerformance appraisal tasks, certainty that they completed tasks successfully, and perceived ease of system use.

Table 5. Usability Pretest and Post-test Survey Outcomes

Variable Description	Overall Mean (n)	Supervisor Status	
		Sup (S)	Non (NS)
Pretest Survey			
I have a clear understanding of the present EVAL/FITREP system. 5: Strongly agree ~ 1: Strongly disagree[1]	4.00 (57)	4.26 [NS]	3.59 [S]
My last EVAL/FITREP was fair/accurate. 5: Strongly agree ~ 1: Strongly disagree	4.11 (57)	4.26 [NS]	3.86 [S]
My last EVAL/FITREP was conducted in a timely manner. 5: Strongly agree ~ 1: Strongly disagree	4.00 (57)	4.23 [NS]	3.64 [S]
I was able to submit my own input at my last EVAL/FITREP. 5: Strongly agree ~ 1: Strongly disagree	4.12 (57)	4.40 [NS]	3.68 [S]
My last advancement/promotion recommendation was fair/accurate. 5: Strongly agree ~ 1: Strongly disagree	4.18 (57)	4.40 [NS]	3.82 [S]
I am satisfied with the present Navy EVAL/FITREP system. 5: Strongly agree ~ 1: Strongly disagree	3.46 (57)	3.57	3.28
The most qualified and deserving Sailors score the highest on their EVALs/FITREPs. 5: Strongly agree ~ 1: Strongly disagree	3.25 (57)	3.43	2.95
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	2.96 (53)	2.94	3.00
How efficient or inefficient do you think the performance management system will be? 5: Very efficient ~ 1: Very inefficient	3.17 (53)	3.28	3.00

Table 5. Usability Pretest and Post-test Survey Outcomes (continued)

Variable Description	Overall Mean (n)	Supervisor Status	
		Sup (S)	Non (NS)
Post-test Survey			
How comfortable or uncomfortable did you feel performing the tasks in the test? 5: Very comfortable ~ 1: Very uncomfortable	3.15 (55)	3.62 ^{NS}	2.38 ^{S}
How certain or uncertain are you that you completed the tasks successfully? 5: Very certain ~ 1: Very uncertain	3.36 (55)	3.85 ^{NS}	2.57 ^{S}
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	3.28 (54)	3.52 ^{NS}	2.90 ^{S}
Overall, how easy or difficult was the system to use? 5: Very easy ~ 1: Very difficult	3.45 (55)	3.56	3.29
Overall, how easy or difficult was the system to understand? 5: Very easy ~ 1: Very difficult	3.53 (55)	3.62	3.38
Overall, how professional or unprofessional did the system appear? 5: Very professional ~ 1: Very unprofessional	4.22 (55)	4.32	4.05
Overall, how efficient or inefficient was the system? 5: Very efficient ~ 1: Very inefficient	3.45 (55)	3.26	3.76
Overall, as you worked through the tasks, did the product become… 5: Much easier to use ~ 1: Much harder to use	4.02 (55)	3.97	4.10
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	3.72 (54)	3.73	3.71
I have a clear understanding of the performance management system. 5: Strongly agree ~ 1: Strongly disagree	3.18 (55)	3.03	3.42

Table 5. Usability Pretest and Post-test Survey Outcomes (continued)

Variable Description	Overall Mean (n)	Supervisor Status	
		Sup (S)	Non (NS)
The performance management system seems fair/accurate. 5: Strongly agree ~ 1: Strongly disagree	3.78 (55)	3.71	3.90
The performance management system allows performance reviews to be conducted in a timely manner. 5: Strongly agree ~ 1: Strongly disagree	3.64 (55)	3.67	3.62
I am satisfied with the test version of the performance management system. 5: Strongly agree ~ 1: Strongly disagree	3.46 (54)	3.33	3.67

[1] The original agreement scale in both survey questionnaire had the opposite endpoints as 1 indicated “strongly agree” and 5 “strongly disagree.” This scale was reversed in the analysis for the presentation convenience.

Note: Superscripts [S], [NS] indicate significantly different estimates at the 0.05 level from *t*-test.

Pre-test perceptions of the FITREP/EVAL system are compared to post-test perceptions of the HPFD/ePerformance appraisal system in ***Table 6***. The analyses indicate supervisors had more frequent significant changes in their pre-test and post-test perceptions. Four items have statistically significant differences for supervisors: clear understanding of system, system fairness/accuracy, system task completion in a timely manner, and system ease of use. All but system ease of use showed a decrease in positive perceptions. Supervisors perceive that the HPFD/ePerformance system may be easier to use than the FITREP/EVAL system. The greatest magnitude of difference between pre-test and post-test was observed in the item assessing a clear understanding of the system. Supervisors may benefit from additional information communicating the purpose of the HPFD/ePerformance system. Fewer significant differences were observed among non-supervisors. Non-supervisors' perceptions of system efficiency increased slightly and the change was statistically significant.

Table 6. Differences between Pre Usability Test and Post Usability Test Survey Outcomes

Item Description	Supervisor			Non-supervisor		
	Pretest (n)	Post-test (n)	Pre-post change[1] (n)	Pretest (n)	Post-test (n)	Pre-post change[1] (n)
Clear understanding of system	4.26 (35)	3.03 (34)	-1.24* (34)	3.59 (22)	3.42 (21)	-0.14 (21)
System fairness/Accuracy	4.26 (35)	3.71 (34)	-0.53* (34)	3.86 (22)	3.90 (21)	0.05 (21)
System task completion in a timely manner	4.23 (35)	3.62 (34)	-0.62* (34)	3.64 (22)	3.67 (21)	0.05 (21)
System satisfaction	3.57 (35)	3.33 (33)	-0.21 (33)	3.28 (22)	3.67 (21)	0.43 (21)
System ease of use	2.94 (32)	3.56 (34)	0.68* (31)	3.00 (21)	3.29 (21)	0.29 (21)
System efficiency	3.28 (32)	3.26 (34)	0.06 (31)	3.00 (21)	3.76 (21)	0.76* (21)

[1] Pre-post changes are computed for cases where both pre-test and post-test items are completed. Therefore, the simple differences between the scores shown above do not necessarily match the pre-post change.

Note. Superscript * indicates that pre-post score change is significant from paired *t*-test.

7 Summary and Conclusions

7.1 Key Findings

Other than problems related to NSIPS (i.e., system timing out and long page loading times) results from testing the HPFD and ePerformance systems indicate that the systems themselves worked well. Overall, the usability survey results suggest no major systematic problems among the groups in the study. However, several minor trends do emerge in the data.

- Non-supervisors had a more difficult time using certain functions in the system, including cutting and pasting, and locating specific buttons.
- Logging onto NSIPS and accessing the HPFD/ePerformance system took longer aboard ship than at shore test sites.
- Supervisors felt more comfortable and confident using the Web-based performance management system.
- Non-supervisors and supervisors gave similar overall ratings of the Web-based performance management system. For supervisors only, the ratings tended to be lower than their ratings of the EVAL/FITREP system.

Because of the differences observed between participant groups, the following are suggestions that may make implementation of the Web-based performance management system operate more smoothly. First, future rounds of testing the Web-based performance management system should focus more heavily on non-supervisors, since non-supervisors appear to have more usability errors. Supporting materials developed during the course of usability testing (i.e., the QRG) should help users navigate within the system. Second, Sailors aboard ship may experience and might need to expect slower system connectivity. All efforts should be made to increase the speed of the system and to decrease the occurrence of the timing out problem. Finally, since supervisors are more likely to be satisfied with the current EVAL/FITREP system, CNPC might consider working with senior enlisted and officers to test and refine the HPFD/ePerformance appraisal system.

7.2 Limitations of Research

The objective of usability testing is typically not to test for group 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. This limitation of usability research limits the generalizability and representativeness of the results.

Relative to other usability testing designs, this study used a reasonably large number of participants - 57 Sailors across the three iterations. The current study doubled the typical number of participants per iteration in order to get an adequate number of supervisory and non-supervisory Sailors for between group significance testing of data from HPFD and ePerformance system usability. The study design does capture a meaningful representation of system users for analysis of the key variable – system usability by supervisory and non-supervisory personnel.

Also, most usability testing study designs call for an iterative approach where revisions to the system under examination and comparisons are made between iterations. Because of the compressed timeframe of this study, the number and types of changes between iterations were limited. While the initial study design called for system modifications between iterations, it became apparent during the first iteration of this study that this aspect of usability testing could not be accommodated.

7.3 Recommendations for Future Research

Results from this study raise several questions that could be addressed through future research. First, while the recommendations from this study are likely to be incorporated in subsequent versions of the Navy HPFD and ePerformance systems, a small follow-up usability study may be needed to confirm that changes made to the system indeed made it a more usable system—fewer NSIPS problems, decreased time spent on completing tasks, and increased satisfaction with the system. While testing the effectiveness of system changes, additional usability tasks could be added to more closely approximate the entire HPFD/ePerformance appraisal process – an e-mail notification that a performance document needed to be created, creating the document, soliciting performance input, completing the input, and routing it through the unit/command for approval.

A subsequent follow-on phase prior to full system implementation would be to conduct a full pilot study where an entire command would complete the performance counseling and appraisal process using the HPFD and ePerformance systems and pilot test results would be compared to a control condition using the current Navy performance counseling/appraisal system. This would enable NPC to determine the relative effectiveness of the HPFD/ePerformance system compared with the status quo – a necessity in adopting new performance counseling and appraisal systems.

8 Closing

This study provides information that can significantly improve the Navy's HPFD and ePerformance system and its implementation. At the very least, making the performance management/appraisal process more efficient should decrease rather than increase the burden on Sailors throughout the Fleet. As mentioned earlier, active participation in the performance appraisal process is a significant component of job satisfaction, organizational commitment, and a likely factor in retention plans. Improving system usability could facilitate a meaningful performance management experience and thus a positive effect on quality of work life, job satisfaction, and possibly Sailor retention.

Results of this study indicate that the majority of usability problems occur early in the system access phase (i.e., logging onto NSIPS) as well as during the course of using the HPFD or ePerformance system document (i.e., system timing out). Previous analyses of focus group data from this study indicates that the less than positive user perceptions of the HPFD and ePerformance appraisal system were primarily attributable to poor system connectivity (Schwerin, Dean, Robbins, Bourne, 2004). Improvements to the NSIPS system that enable consistent and reliable access to the HPFD and ePerformance documents are likely to greatly enhance the system usability and user satisfaction.

While usability testing is just one phase of the system design, testing, and implementation process, it is crucial that system refinement and improvements be made at this point rather than waiting until the implementation phase. Initial system training and subsequent retraining can potentially be frustrating to users and could lead to errors in the performance appraisal process that could generate feelings of alienation and disenfranchisement.

Finally, beyond the issue of system usability is the issue of ensuring the new performance management/appraisal system is better and has no differential negative effect on constituent groups of Navy personnel (e.g., race, gender, ethnicity, warfare community, etc). The Code of Federal Regulations (1978) and professional associations that guide the ethical implementation of personnel selection, testing, and appraisal (AERA, 1985; SIOP, 2003), require large-scale pilot studies comparing a new system to the current system prior to implementation. Large-scale pilot studies should be designed to examine 1) the effectiveness of the new system on the criterion measure (e.g., comparing the new system for evaluating workplace performance to the current system and possibly a third, independent measure of workplace performance and 2) the existence of differential negative effects of the new system on constituent groups of Navy personnel (i.e., disparate treatment among groups). These subsequent phases to system implementation need to be considered and conducted for the successful implementation of the HPFD and ePerformance appraisal system for the Navy's active duty and reserve workforce.

9 References

American Educational Research Association (AERA; 1985). *Standards for Educational and Psychological Testing*. Washington, DC: AERA.

Bevan, N., & Macleod, M. (1994). Usability measurement in context. *Behavior and Information Technology, 13*, 132-145.

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* (Institute Report #429). Tampa, FL: Personnel Decisions Research Institute, Inc.

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

Bureau of Naval Personnel Instruction (BUPERSINST) 1600.10 (1995). *Navy performance evaluation and counseling system* (PERS-32). Washington, DC: Bureau of Naval Personnel.

Clark, V.E. (December 3, 2001). *Opening remarks to the chief of naval operations (CNO) leadership summit.* Monterey, CA: Naval Postgraduate School.

Code of Federal Regulations (1978). *Uniform Guidelines on Employee Selection Procedures (Part 60-3).* Title 41: Public Contracts and Property Management, Chapter 60: Office of Federal Contract Compliance Programs, Equal Employment Opportunity, Department of Labor. Washington, DC: Government Printing Office (GPO).

Dailey, 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.

Department of Defense Instruction (DODINST) 5000.2 (2003). *Operation of the Defense Acquisition System*. Washington, DC: Department of Defense (USD AT&L).

Dolan, W.R., & Dumas, J.S. (1999). A flexible approach to third-party usability. *Communications of the ACM, 42*(5), 83-85.

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.

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 non-supervisory jobs in the U.S. Navy* (Institute Report #415). Minneapolis, MN: Personnel Decisions Research Institutes, Inc.

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* (Institute Report #453). Minneapolis, MN: Personnel Decisions Research Institutes, Inc.

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

Krueger R. (2003). *Focus groups: A practical guide for applied research* (3rd ed.). Thousand Oaks, California: Sage Publications.

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.

Morgan D. (1997*). Focus groups as qualitative research* (2nd ed.). Thousand Oaks, California: Sage Publications.

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

Norman, K.L, Friedman, Z., Norman, K., and Stevenson, R. (2000). Navigational Issues in the Design of On-Line Self-Administered Questionnaires. Human-Computer Interaction Laboratory, University of Maryland. *Unpublished Manuscript.* Accessed from ftp://ftp.cs.umd.edu/pub/hcil/Reports-Abstracts-Bibliography/2000-22html/2000-22.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., Robbins, K.M., & Bourne, M.J. (2004). *Usability testing of the U.S. Navy's performance management system: Technical report #1*. Research Triangle Park, NC: RTI.

Society for Industrial and Organizational Psychology, Inc. (2003). *Principles for the Validation and Use of Personnel Selection Procedures*. Bowling Green, OH: SIOP.

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.