

Assessing Usability of a Post-Mission Reporting Technology

A Novel Usability Questionnaire in Practice

Mitchell J. Tindall¹(✉) and Beth F. Wheeler Atkinson²

¹ StraCon Services Group, LLC, Orlando, FL, USA

mitchell.tindall.ctr@navy.mil

² Orlando, FL Naval Air Warfare Center Training Systems Division,

Orlando, FL, USA

beth.atkinson@navy.mil

Abstract. Usability evaluation has received extensive attention in both academic and applied arenas. Despite this, there have been few formal attempts to integrate past research and best practices in an effort to develop a newly updated and adaptable approach. This poster provides an overview of the types of results yielded by a novel usability assessment approach (i.e., Experienced-based Questionnaire for Usability Assessments Targeting Elaborations [EQUATE]) when applied to a post mission reporting tool. The goal of this study was to develop software to automate performance tracking for anti-submarine aircraft, digitize performance and training information, and automate the display of post mission summaries. Although some of these technologies exist, the prototype tested during this research was the first, of which the authors are aware, to provide a single point of access for data entry, analysis and reporting. Due to the potential benefits across a variety of naval aviation platforms, the program's usability goals focused on identifying means to optimize the tool by gathering novice user feedback. Traditional methods for end-user feedback have tended to focus on user performance and satisfaction, rather than providing prescriptive inputs to identifying and rectifying issues. The results of this study provided usability input for post mission reporting, as well as identified and narrowed the heuristic dimensions used for final validation.

Keywords: Usability · Heuristic evaluation · GUIs

1 Introduction

Designing Graphic User Interfaces (GUIs) for complex applications such as those used to support naval operational and training systems is a challenging endeavor. The challenges arise from the balance between appeasing the functional needs of the

The views expressed herein are those of the authors and do not necessarily reflect the official position of the Department of Defense, its components, or the organizations with which the individuals are affiliated.

end-user population with ensuring the usability of the system. Despite the challenges, the goals should always be to develop systems that are accessible, easy to use, make task completion more efficient, reduce the cognitive workload on the end user, and generally satisfy the user. Even the most seasoned computer programmers cannot develop such systems without early, frequent and prolonged end-user feedback. Such feedback has traditionally been solicited through a process of usability testing. While no single definition of usability exists in the literature, it is generally defined as the quantifiable characteristics of a system that provide information to developers about the ease of use and usefulness of the system [1, 2, 5]. Literature regarding usability evaluations suggests they are necessary for developing usable interactive systems [3, 4]. While the challenges of conducting usability testing for Navy systems can be unique (e.g., domain specificity, complex systems of systems), they are not immune to the benefits of such evaluations. The following report details an initial usability evaluation of the Post Mission Assessment for Tactical Training and Trend Analysis (PMATT-TA) Increment 1 web-based application using a novel, heuristic-based measure called the Experience-based Questionnaire for Usability Assessments Targeting Elaborations (EQUATE).

1.1 PMATT-TA System Purpose and Goals

PMATT-TA Increment 1 is an online web-based application and database designed for tracking important data points from operational and training events for the maritime patrol community. The post mission data captured includes a variety of missions (e.g., mission type, communications) and contextual information (e.g., weather). The ultimate goal for capturing this data is to facilitate debriefs with aircrews to identify strengths and areas for improvement. By centrally storing post mission information, PMATT-TA also supports data calls related to trends (e.g., number of events a crew has participated in). Additionally, data increments of PMATT-TA will address a need for a digitally-based system that streamlines and automates the process of data collection, analyses, and feedback. As a result, the system benefits a range of users (e.g., aircrew, instructors, squadron, or group leaders). While the demand for PMATT-TA within the Navy is apparent, it has not yet been subject to formal usability analysis. Usability analyses must be conducted to be confident that the system achieves its goals, reduces post mission reporting time and is operable by end-users.

2 Data Collection

2.1 Recruitment

Participants ($N = 9$) evaluated the PMATT-TA Increment 1, web-based application in a laboratory setting with a desktop computer. After attaining an informed consent from qualified participants, researchers provided a brief explanation of tasks and measures as they moved through the protocol.

2.2 Measure

After completing the PMATT-TA tasks, each participant was given the EQUATE. It was developed based on an extensive review of the extant literature regarding system design and heuristic evaluation [6]. The review identified a number of items (i.e., 250) that qualified as design guidance. Through further testing eight heuristic dimensions (i.e., *Error Handling & Feedback*, *Graphic Design & Aesthetics*, *User Interaction Control*, *Memorability & Cognitive Facilitation*, *User Efficiency*, *Learnability*, *Consistency*, and *Help* were included within the EQUATE). Validation in the form of internal consistency and discriminate validity was ongoing at the time of this study but preliminary analysis has demonstrated sound psychometric properties (i.e., average internal consistency for all dimensions $< .85$).

3 Results

The quantitative analysis is intended to provide a general overview of the usability of the PMATT-TA system and a localized evaluation of the usability of system and system components across the heuristic categories previously mentioned (e.g., *Learnability*, *Help*, and *Consistency*).

The analysis revealed both positive and negative elements within the PMATT-TA system (see Table 1). Usability was assessed on a 5-point scale (i.e., 1 = *Strongly disagree*, 5 = *Strongly agree*, 0 = *Not applicable*). Items on the EQUATE are framed in both positive and negative terms (e.g., *The design provided a pleasant experience* and *There was too much clutter on the display*). Negatively framed items were reverse coded prior to analysis. Higher averages imply better usability while lower averages imply potential usability issues exist.

The average visibility across all items of the EQUATE was acceptable ($M = 3.41$, $SD = 0.24$). While this average appears to indicate the system demonstrates adequate usability from a global perspective, a more detailed evaluation of EQUATE items, dimensions, and free-responses is necessary to validate and elaborate on this assertion. Two of the heuristic categories exhibited an average score below 3.0, indicating the need for redesign, adjustment or enhancement: *Learnability* ($M = 2.89$, $SD = 0.56$) and *Help* ($M = 2.81$, $SD = 0.93$). This demonstrates that participants (i.e., usability evaluators) felt they needed more or better help to learn the system¹. The remaining dimensions all demonstrated adequate or above adequate average scores (i.e., ≤ 3) (see Table 1 for descriptive statistics).

The opportunity for participants to articulate system issues in a free-response format was exercised extensively. The vast array of free responses required the research team to qualitatively (i.e., summarize for system developer feedback) and quantitatively (i.e., coding into respective heuristic dimensions) analyze the data provided. The

¹ The participants in this study lacked domain specific knowledge on how to complete a post mission report (i.e., not military participants) likely influencing their ratings of learnability. For this reason, *Learnability* for the system should be re-evaluated in future iterations of testing with end users to further evaluate this factor.

Table 1. Descriptive statistics for Evaluator responses on the EQUATE ($N = 9$)

Dimension	Range		M	SD
Graphic Design and Aesthetics	3.34	3.95	3.681	.1968
Error Handling and Feedback	3.19	4.19	3.506	.3131
User Interaction Control	3.20	4.00	3.502	.3161
Memorability and Cognitive Facilitation	2.48	4.15	3.280	.5122
User Efficiency	3.16	4.21	3.687	.4324
Learnability	2.10	3.88	2.897	.5646
Consistency	2.93	4.50	3.875	.5733
Help	1.17	3.91	2.816	.9398
Overall Usability	3.13	3.90	3.409	.2486

Note. The potential range of responses for all heuristic dimensions was 1 to 5 (i.e., 1 = *Strongly disagree*, 5 = *Strongly agree*, 0 = *Not applicable*)

quantitative analyses validated EQUATE survey items; *Learnability* and *Help* were the most frequently mentioned dimensions. In addition to validating survey responses, free-responses lent participants the opportunity to elaborate on the specifics of usability issues through severity ratings of those issues and suggesting ideas for fixing them. Specific issues were identified in subject free-responses that would not have been discovered with an analysis of only survey results.

For example, one field for inputting a unit's location was not working properly (i.e., would not accept letters, numbers or any combination of the two). Not only did this prevent participants from entering in the necessary information, it prevented them from adequately completing the overall task. Each of the nine subjects identified this as a *highly severe* issue. The information was communicated to the development team and addressed appropriately.

Aside from identifying obvious fixes, free responses embedded in the EQUATE survey allowed participants to elaborate on issues. For example, while less than half the participants reported subpar *User Efficiency* in survey items, an examination of their free-responses yielded important information for system usability. They indicated that the procedure for creating a *training event* in PMATT-TA was not sufficiently clear. They described this issue as *severe* and offered suggestions for fixing it (i.e., making the *New Event* button on the home page more conspicuous by bolding, adding color or increasing its size). Because creating a *New Event* is one of the system's main purposes, this information was vital to the system development team to enhance usability. In sum, though tedious to analyze and interpret, the opportunity for subject free-response proved essential for improving and especially fixing the PMATT-TA system.

4 Discussion

It is the goal of any armed forces to maintain a competitive advantage over the enemy by having the most advanced capabilities. Such capabilities are the result of constantly evolving hardware and software technologies. Unfortunately, developing these

technologies is only one half of the equation. If they cannot be used effectively, their development has limited impact or desired effect. As a result, inexpensive, efficient, adaptable and easily analyzable systems for evaluating the usability are essential. PMATT-TA was the first system to be evaluated using a novel usability assessment method, which was developed based on an extensive review of literature and best practices. This initial analysis indicates the EQUATE provides a reasonably comprehensive, inexpensive, efficient, adaptable, and easy to analyze and interpret method for capturing end user feedback. Because of the results of this analysis, the development team benefited from the identification of critical issues prior to product development. While both the PMATT-TA and the EQUATE are still receiving research attention for validation and general improvement, preliminary analyses demonstrate the utility of each system in advancing the capability of the U.S. Navy.

Acknowledgements. This research was sponsored by the NAVAIR Section 219 and PMA-205 Air Warfare Training Development programs. We wish to thank interns who facilitated data collection and analysis, and colleagues who provided input throughout this process. The views expressed in this paper are those of the authors and do not represent the official views of the organizations with which they are affiliated.

References

1. Bowman, D.A., Gabbard, J.L., Hix, D.: A survey of usability evaluation in virtual environments: classification and comparison of methods. *Presence Teleoperators Virtual Environ.* **11** (4), 404–424 (2002)
2. Hix, D., Hartson, H.R.: *Developing User Interfaces Ensuring Usability Through Product and Process*. John Wiley, NJ (1993)
3. Shneiderman, B.: *Designing User Interface Strategies for Effective Human-Computer Interaction*. Addison-Wesley Reading, MA (1998)
4. Nielsen, J.: *Usability Engineering*. Academic Press, New York (1993)
5. Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., Carey, T.: *Human-Computer Interaction*. Addison-Wesley Longman Ltd, London (1994)
6. Atkinson, B., Tindall, M., Igel, G.: Validated Usability Heuristics: Defining Categories and Design Guidance. Extended poster abstract submitted to HCI International (2015)