User-Centered Design in a Large-Scale Naval Ship Design Program: Usability Testing of Complex Military Systems—DDG 1000

Vince Quintana, Robert A. Howells, and Lawrence Hettinger

Abstract

The US Navy is currently developing a new class of surface combatant ships referred to as DDG 1000. Unlike past ship designs, DDG 1000 is explicitly focused on the execution of a user-centered approach—one that captures the requirements, capabilities, and limitations of sailors expected to operate the ship. Additionally, DDG 1000 presents a number of significant human factors and challenges related to its greatly reduced crew size (compared with similar legacy ships) and high reliance on automated systems. Usability testing and usability assessments (UT and UAs) for the purpose of validation of human systems integration (HSI) principles, and the validation of the DDG 1000 Sailor Systems Specification (S3), play a key role in the DDG 1000 Human-Centered Design (HCD). Through the use of summative and formative design analysis and design verification events, UT/UA is critical in identifying and rectifying issues early in the design process, and verifying system functionality later in the design process. For the DDG 1000 program, the Human Systems Integration—Design Verification Integrated Product Team (HSI-DV-IPT) has also combined its efforts with both the DDG 1000 Safety and Training Cross Product Teams to realize the economies of scale associated with conducting these tests to obtain data for the needs of all three HCD disciplines. Through the course of the DDG 1000 detailed design phase, UTs and UAs have become a defined procedural process, which is governed by both the DDG 1000 Human Systems Integration Plan and as an extension of the total testing plan for the DDG 1000 ship. Almost all DDG 1000 HSI-DV-IPT evolutions combine the efforts of system engineering and design teams with HSI engineers and fleet users to ensure that the combined test output product has received a "cut" from the entire chain of influence from concept to end user. To date, over 30 user interaction and test events employing 1,100 users have been conducted to verify the utility of aspects of the DDG 1000 design, from the most sophisticated to the most mundane systems and processes. In addition, a series of tools that includes 3D visualization aids for modeling and simulation, and weighted assessment systems like NASA-TLX have been included in the DDG 1000 HSI test engineers' tool bag to provide the best of breed solutions to HCD-oriented testing. The test results obtained to date provided valuable insight into the validity of the DDG 1000 HCD and provided feedback that led to optimization of both crew and ship as its HCD objectives intended, in addition to providing cost-effective and timely feedback into the DDG 1000 design.

Purpose

The purpose of this report is to review the tools, processes, and procedures utilized by the DDG 1000 Human Systems Integration—Design Verification Integrated Product Team (HSI-DV-IPT), examine the challenges they faced and the results of their efforts, while extracting general lessons learned for possible application elsewhere in the future. It is hoped that the reader will come away with a better understanding of the potential roadblocks to practical application of human systems integration (HSI) principles and techniques while attempting to mitigate HSI issues identified in a complex socio-technical design.

Introduction

The US Navy is currently engaged in the development of a future surface combatant referred to as DDG 1000. DDG 1000 will be a class of ships that will differ greatly from its predecessors, with respect to both its design characteristics (including crew size and supporting hardware and software systems) and the systems engineering processes that are driving its development.

DDG 1000 will have a greatly reduced crew size relative to comparable legacy ships (approximately 140 full-time crew with air detachment on DDG 1000, as compared with approximately 350 on comparable legacy ships) and will incorporate a number of new ship-based systems that will rely heavily on automation, in part to compensate for the reduced crew size. DDG 1000 is also the first ship-building program in the history of the US Navy to adopt a user-centered design approach explicitly. The DDG 1000 HSI Cross-Product Team, composed of human factors and ergonomics specialists from industry and government, is an integral component of the overall systems engineering effort.

Human factors and ergonomics usability testing (UT) have taken on a significant role in the DDG 1000 program as one of the key elements in the overall user-centered design approach. Within the DDG 1000 program, there is an Integrated Product Team whose primary focus is the conduct of usability assessments (UAs).

Usability is the measure of the quality of a user's experience when interacting with a product or system and can be assessed using predefined objectives and metrics associated with directly observable behaviors such as reaction time, error rates, number of operations required to complete a task, etc. It can also be assessed using measurements of cognitive processes driven by workload and situation awareness.

UAs are composed of a set of techniques for ensuring that the intended users of a system can carry out intended tasks efficiently, effectively, and satisfactorily. They are usually carried out before system release so that any significant issues identified can be addressed; however, they can be carried out at essentially all stages of the design process, although the appropriate technique varies according to the level of systems maturity. For the purposes of the DDG 1000 program, HSI UT is being conducted using three testing activity foci (design analysis (DA), design verification (DV), and HCI testing events) across two states (formative and summative) of the product maturity domain (see **Figure 1**).

Formative testing refers to testing that is performed while the product is being planned, designed, and initially developed. The most general goal is to detect and remedy gross problems early in the design process. This is also known as testing within the immature product domain. Formative testing is an important aspect of the DDG 1000 UA process. Summative testing refers

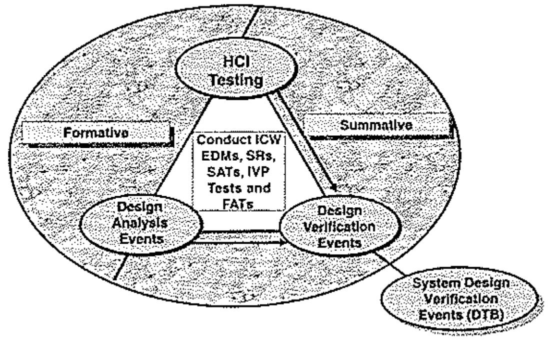

Figure 1: *The Human Systems Integration—Design Verification Integrated Product Team has three foci in two product maturity state domains*

to testing that is usually performed after formative testing (and after resulting design changes have been made). The product is generally at a final or near-final form when summative tests are performed, and the testing involves more subtle issues than formative testing. This is also known as testing within the mature product domain. Within the DDG 1000 program, summative tests will primarily take the form of DV as opposed to DA or design guidance events.

UAs are part of the process of usability engineering. UAs include a range of methods for having users try out a system. In a typical UA, users perform a variety of tasks with a prototype (or other system) while observers collect data and/or record notes on what each user does and says. Typical assessments are conducted with one user at a time or two users working together. Concept of Operations Exercises (COOPEXs) may involve entire teams of users performing representative tasks with prototype systems. Assessments may include collecting data on the paths users take to execute tasks, the errors they make, when and where they are confused or frustrated, how fast they perform a task, whether they succeed in performing the task, the level of physical or cognitive workload involved in performing a task, and how satisfied they are with the experience. The goal of most UAs is to uncover any problems that users may encounter so that those problems can be fixed. UAs can also be focused on specific subtasks or subroutines within a system.

Given the above requirements, the primary concentration of the DDG 1000 HSI-DV team has been to develop a test program with a focus on two major areas:

DA events—DA UT events are intended to provide useful feedback, as early in the system design process as possible, on the usability of key DDG 1000 human–computer and human–machine systems. The results of these UT events are used to generate new design requirements, or changes to existing design requirements, for the express purpose of enhancing the usability, safety, and efficiency of the use of key DDG 1000 systems.

DV events—DV UT events occur relatively later in the design cycle, and are intended to provide empirical verification that the design of key DDG 1000 systems satisfies user-centered design requirements established earlier in the design effort.

Within these two broad areas, UT focuses on the reduction of risk associated with the inadequate usability of human–computer and human–machine systems aboard the ship. Such risk, of course, manifests itself in terms of frequent errors, performance that is insufficiently rapid and/or precise, and the generation of workload that is unacceptably high (or low) as well as unsatisfactory levels of situation awareness. Deficits in any of these areas represent key risks to the overall DDG 1000 crew concept, and if left unidentified (by, for instance, not conducting sufficient front-end DA) could significantly interfere with the overall performance of the crew and ship.

In the detailed design (DD) phase, assessments are being conducted on products that have advanced beyond a purely formative form. Some products have been subjected to at least some preliminary design evaluation and testing (achieving borderline maturity) and may thus be partially formative and summative. Some products are quite mature and in near-final form; hence, tests on these systems will be considered as summative tests. The DV-IPT will focus its activities in three ways on product designs distributed between the formative and summative product maturity state domains. These activities will be termed HCI usability events, DA events, and DV events. The HCI and DA events will be conducted on systems with varying but significant degrees of design immaturity, while DV events will be conducted on systems that are relatively mature. These three classes of assessment activity will be conducted in accord with product teams, system reviews, tests defined in the integrated verification plan, system acceptance tests, and final acceptance tests.

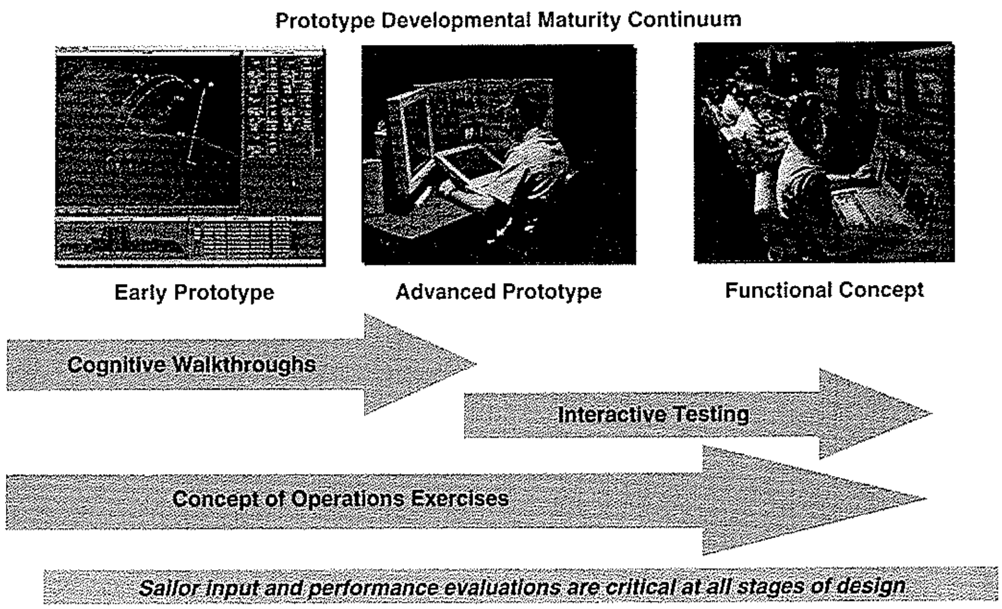

Figure 2: *Usability testing approaches as a function of product maturity*

COOPEX EVENTS, COGNITIVE WALKTHROUGHS, AND INTERACTIVE EVALUATIONS

DD phase UAs require the application of a number of different evaluation techniques, primarily due to variations in the level of developmental maturity of the systems under investigation, as well as the nature of the design questions to be addressed. These evaluation techniques (as shown in **Figure 2**) include Cognitive Walkthroughs, COOPEXs, and Interactive Evaluations.

As an example of the types of issues examined to date, several DA events have been carried out to examine the ease, efficiency, and effectiveness of the use of the human–computer interfaces being designed for use within the DDG 1000 Ship Mission Center—the tactical nerve center of the ship. Other DA events have focused on the design and layout of equipment on the ship's Bridge, Galley, and Damage Control facilities among others. In each case, results from these UT events have resulted in significant modifications to the design of ship systems, enhancing their overall usability and effectiveness.

A variety of analytic techniques are used to support DA UT and UA events. These range from "cognitive walkthroughs" (cf. Dumas and Redish 1995) of very early conceptual depictions and representations of candidate systems to the use of fully functional prototypes for hands-on Human-In-The-Loop (HITL) UT. The purpose of these events is to expose representative users (generally active-duty US Navy fleet personnel with expertise matched to the system under consideration) to various aspects of the design when possible, and also to expose the design to examination by HSI engineers to head off potential HSI problems of the design in its earliest stages. The results of UTs and UAs conducted to date have also resulted in significant modifications and improvements to the DDG 1000 design.

UTs/UAs and Methods Used within the DDG 1000 Program

FULL-DIMENSIONAL MOCK-UPS

For several of the formative usability assessments, the creation of a full-dimensional mock-up was used, each built to the design specifications defined by the latest architectural dimensions (**Figure 3** is typical). The primary purpose of these mock-ups was to evaluate the utility of the area in question to support the watchstations and functions contained within those areas for any indications of design-imposed human performance limitations (i.e., line of site). In addition, end users were

Figure 3: *Helicopter control station mock-up*

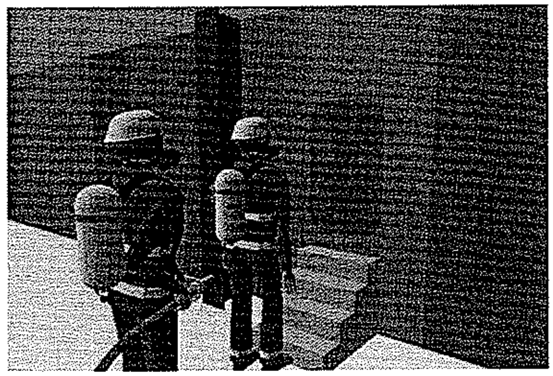

Figure 4: *Simulation modeling for human factors*

brought to the test to provide a subjective analysis of the design in question to ensure that the designers and engineers had in fact "gotten it right" from an end-user perspective. It also provided the systems and space design engineers with an opportunity to obtain data from experienced end users on the optimum placement of equipment, displays, and controls before completion of the watchstation designs.

The advantage of this type of assessment comes from the ability to make the "boxes" of which it is made rapidly reconfigurable. Operators were polled on their best assessment of the location of various devices, and the test conductors were literally able to reconfigure the space "on the fly." With weighting and discussion, this would lead to the best of breed consensus-derived configuration whose performance could be further checked with a DV event.

MODELING AND SIMULATION (MISSILE LAUNCHER COMPARTMENT)

For modeling and simulation needs, two types of analysis tools were used by the DDG 1000 HSI-DV-IPT (physics-based and dimensional-based modeling). The use of modeling for dimensional-based analysis is primarily a DA tool and was in many cases used to catch a quick or cursory advisory look at design products. Several system and component engineers requested the use of DV support to resolve the use of non-standard solutions in cases where those solutions may have impacted crew performance. An example of this was the review conducted of the peripheral vertical launch system entrance door, which was a non-standard design but required review to ensure the door was usable by the crew, especially in emergency situations. A volumetric/dimensional model was created of the notional design, and 3D modeling was used to verify that the door in question was in fact usable by a 95 percentile male in full Damage Control Ensemble (see **Figure 4**).

The use of 3D modeling is an effective tool and provides for the ability to make assessments without actually having to "cut steel." Properly utilized, it is a great cost and time saver, and provides for rapid feedback through the use of almost instantaneous parameter change (electronic bulkheads are easier to move). It is important to note here, however, that the most important component of this method of testing is to ensure that all components of this solution set, including software, tools, and device models, are carefully and completely validated through a recognized validation and verification process.

PERFORMANCE ANALYSIS (LINE-OF-SIGHT STUDY)

Another test tool that has been used is the execution of performance analysis studies conducted to quantify performance using more traditional mathematical and engineering approaches (**Figure 5**). This method is especially

Figure 5: *Operator line-of-sight study for aircraft operations*

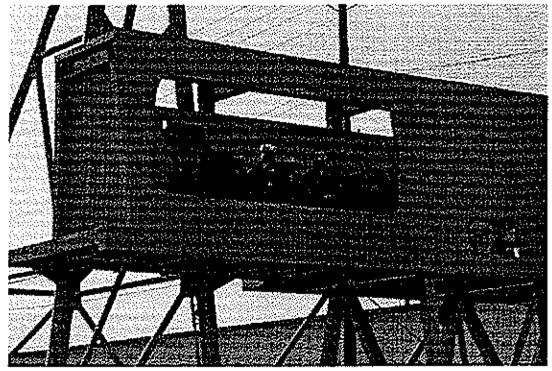

Figure 6: *Human-In-The-Loop testing of the DDG 1000 refueling station*

important when determining precision performance metrics required to support either design decisions or precise estimations of the crew's abilities, especially critical tasks. As an example, a performance analysis test was required to determine the field of view that a given design would provide crew members to support a specific watchstation. Combining design data with known human ergonomic distribution data enabled the test engineer to determine what capabilities of field of vision the operator must have to support the accomplishment of a specific task, which in this case was to observe the flight deck during aircraft recovery and takeoff. The results of this particular case indicated that a change of the operators' location was required, and retesting was able to conclude that the new position would provide much better operator performance.

HITL—CONFIRMATION OF MANNING ESTIMATES (POCKETIZED REFUELING STATION)

A number of the DDG 1000 HSI-DV-IPT UTs were driven by a need to confirm that estimates of design performance to enhance crew performance and reduction in crewing were viable. A significant portion of the HSI effort on DDG 1000 has been to reduce the crew size to support minimum crewing. This, however, cannot be successful if the remaining crew members are incapable of safely supporting sustainment evolutions essential to supporting a ship's operations. An example of this is the HITL UA that was conducted to confirm that the ship could refuel using a new configuration "pocketized fueling station" that is located behind a large sliding door inside the skin of the ship (**Figure 6**). With the cooperation of the In Service Engineering Agency (ISEA), a fully functional mock-up of the station was built, and refueling evolutions were conducted. This was done using only those number of personnel assigned to watchstations planned to conduct that operation in accordance with the Proposed Ships Manning Document. The tests were quite successful and some safety improvements were recommended to make the evolution even safer, and in fact proved that the evolution could be performed with fewer crew members in an emergency. It also provided the design engineers with an opportunity to conduct "overnight" reconfiguration of the space to find the best possible configuration of equipment and controls.

HITL—PROCEDURAL SEQUENCE PERFORMANCE (ADVANCED GUN SYSTEM)

As discussed by Kaber and Endsley (2001), one of the most significant human performance problems associated with the use of extensive automation is the creation of the OOTL or "Out Of The Loop" problem discussed in their research. For the DDG 1000 HSI team, one of our primary concerns has also been the ability of the crew to perform certain emergent tasks when the ship is damaged, responding to secondary operational methods when automation fails, or simply to support those "cats and dogs" evolutions that are such an important part of everyday sustainment of the ship and that automation cannot perform (there are some things people have to do for the ship . . . like unclogging the toilets).

Several of these tests have been conducted to date, with considerable emphasis being placed on ensuring that these tasks can be accomplished with the tools and procedures at hand, that they can be performed safely, and that the crew is provided with the necessary training to perform these tasks if they are called upon to do so. Given the maturity of various aspects of the design, much of this work to date has been with the ship's weapons systems and has led to early identification of potential problems, as well as provided positive indications that these systems have been, for the most part, Human-Centered Design (HCD) optimized.

For example, in **Figure 7**, two experienced fleet sailors are shown executing an emergency recovery procedure for the Advanced Gun System magazine to demonstrate that the ammunition pallets could be relocated by the crew in the magazine, in case of system failure. This effort was a cooperative test and analysis that combined the efforts of the ISEA, HSI engineers, safety engineers, the DDG 1000 training Cross Product Team, the original equipment manufacturer (OEM), and the prototype control authority.

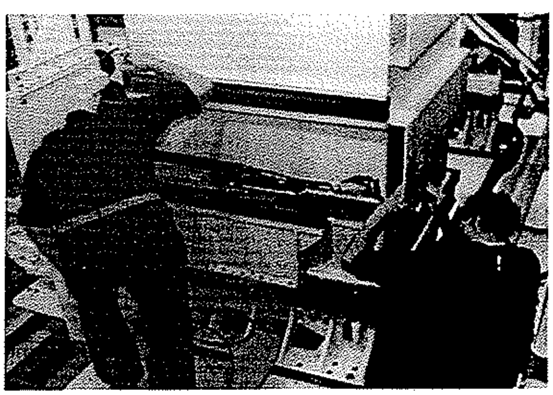

Figure 7: *Fleet test participants conducting emergency recovery procedures in the Advanced Gun System magazine*

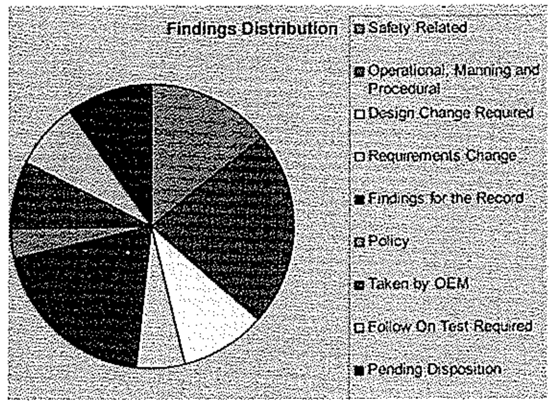

Figure 8: *Summary of usability test findings (relative distribution)*

Lessons Learned from UTs/UAs and Methods Used within the DDG 1000 Program

We can summarize our lessons learned into two categories: process lessons learned and design lessons learned. The design lessons are derived from the test outcomes and are summarized in **Figure 8**. It is important to note that these outcomes have a very specific ship-building taxonomy assigned to them and that those labels are associated with their adjudication. The individual items, however, have very specific HSI, safety, and training labels assigned to them and are maintained as such in the DDG 1000 HSI test outcomes Significant Findings Database. Post test, they are all examined for root cause and worked in accordance with the DDG 1000 Human Systems Integration Plan-defined process and fed back into the design using prescribed engineering processes. They are not all easily mitigated as the usual compromises that need to be negotiated (space, weight, cost, etc.) exist with the DDG 1000 program as with all ship-building programs. However, by adding a scientific process to this evolution, it enables the HSI engineers to provide more credible value to their arguments and recommendations. Even more importantly, by combining these efforts with other Cross Product Teams (Training and Safety), it provides a greater return on investment from the events as the data gathered provide feedback on several aspects of the design simultaneously and reinforce the priority of some outcomes.

The process lessons learned are very specific overall to process, and the logistics of test execution. For DDG 1000, the HSI UT program is an ISO-compliant, Work Instruction-prescribed process. Each step of the process, from candidate test selection and prioritization to outcomes adjudication and prioritization, is carefully defined and monitored for both completeness and quality. They are also governed by fiscal realities as not everything can be tested; hence, a reasonable means approach to prioritization is a must.

SELECTING TEST CANDIDATES

Candidates were chosen using a grading system based on the following priorities:

Design Uncertainty – Design uncertainty refers to those design items that either the systems engineers or the HSI team deems as having some effect on the crewing plans and models. They are characterized within the Manning Uncertainty Issues List (MUIL) and are then assigned a degree of relationship between a proposed test event candidate and an MUIL item or items. If a candidate event directly addresses an MUIL Category 1 item, then it has a higher degree of applicability than if it addresses an MUIL Category 3 item. MUIL item category values are assumed to reflect HSI management and program priorities.

Manning KPP Requirement or Issue – The manning KPP value is a set target that includes both the crew and the air detachment. When a design does not meet requirements, it is possible that the manning KPP will be at risk if more crew members must be added to resolve the design problem. A candidate event that can detect/prevent a design problem that requires a large manning increase will have a high priority value, while an event that involves smaller potential manning increases will be assigned a lower priority value.

Human Factors Engineering Considerations – A system may fail to meet requirements due to a failure of hardware components, human components, or both. This criterion refers to all aspects of system performance or functions that are related to human components of the system. This includes human performance considerations such as operator workload, speed, accuracy, and situation awareness and traditional human factors and ergonomics considerations such as lighting, control arrangement, workspace layout, operator positioning, and dimensional compatibility of workspaces and clearances.

System Operational Safety – This criterion refers to the probability that one or more crew members may be killed or injured if a system does not perform as required. As the probability and number of injuries (which can be detected via a candidate event) increases, the priority of the candidate event increases.

HSI Requirements Coverage – This criterion refers to the spread of assessment events across the set of HSI requirements. It is important that all HSI requirements be verified. If a few requirements are covered by most of the events, but others are left unverified, then there is a problem. Events that are useful for assuring that all requirements will be verified will be considered to have high priority.

SELECTING A TEST METHOD

Each test is then run through a series of filters beginning with **Table 1**, to determine an execution method, location, and opportunity, so that the test can be set in motion. The test methods are prescribed by DDG 1000 requirements documents, and the Total Ship Integrated Test program and schedule. Every attempt is made to use "targets of opportunity," in other words tests that are already scheduled, to leverage costs and availability. The HSI-DV-IPT has "piggy-backed" as many test opportunities as it can, given schedule and resources.

Conclusion

HSI UT is a significant and vital element of any HCD effort. A comprehensive and

TABLE 1: Human System Integration Usability Testing/Usability Assessment Test Method Selection Options

Method	Method Description
Analysis	The use of model and simulations, graphs, tables, and a variety of other analytical tools and methods to analyze the ability of a design to meet specific requirements
Demonstration	An actual system or near-final form prototype is operated against its requirements in order to confirm compliance
Inspection	The physical observation of an installed system to confirm that it conforms to applicable standards requirements (usually performed to assure that the correct equipment has been delivered, properly installed, and is complete)
Test	A system is provided with a planned set of inputs to confirm that corresponding outputs conform to requirements Inputs are generally selected to exercise required system functions across required performance ranges and to confirm the handling of faults and critical safety functions

methodological plan, specific to the program under design, is however required for it to be effective and the program-specific methodology and processes must be agreed upon and implemented early so that early entry economic and HCD-influenced returns can be realized to greatest opportunity. This must include everything from selecting test items and objectives to adjudicating the outcomes. A synergy of (Safety, Training, OEMs, etc.) effort to maximize the tests to provide relevant feedback to both designers and system engineers as early in the design as possible cannot be overemphasized as change cost growth (a key program driver) expands almost explosively as the program matures and makes HCD change implementation even more difficult. There are many tools and methods for actually executing the tests and assessments, and the efficacy of each needs to be defined on a test-by-test requirements basis, with flexibility being the operative word.

Lastly, the test methods selected must be flexible enough to examine, even at small sample rates, all aspects of the human components of the design.

References

Dumas, J.S. and J.C. Redish, *A Practical Guide to Usability Testing*, Intellect Publishers. Exeter, UK, 1995.

Kaber, D. B. and M. R. Endsley, "Theoretical Issues in Ergonomics Science. The Effects of Level of Automation and Adaptive Automation on Human Performance, Situation Awareness and Workload in a Dynamic Control Task," http://www.satechnologies.com/publications/list.php?topic=3, 2001.

Author Biographies

Vince Quintana of Bath Iron Works, is the principal engineer for Human Systems Integration testing for the DDG 1000 program. His background includes patented development work for next generation human borne computing and display systems, and Augmented Reality applications. He has a BS-WED from Southern Illinois University and is a retired US Navy Senior Chief Petty Officer.

Robert A. Howells, Ph.D., is an Engineering Projects Leader and Human Factors Engineer, Northrop Grumman Ship Systems, supporting DDG(1000) Program HSI Design, Design Analysis, and Design Verification activities.

Lawrence Hettinger received his Ph.D. in Experimental Psychology from the Ohio State University in 1987. He has conducted research on human performance for all branches of the military and NASA, and is currently Technical Director for Human-Systems Integration for the US Navy's Future Surface Combatant Program.