

A Practical Guide to the CIF: Usability Measurements

Mary Theofanos | National Institute of Standards and Technology | mary.theofanos@nist.gov

Brian Stanton | National Institute of Standards and Technology | brian.stanton@nist.gov

Nigel Bevan | Professional Usability Services | nigel@nigelbevan.com

A \$46 MILLION COMPUTER SYSTEM installed at the US General Services Administration (GSA) regional offices in Denver and Philadelphia in 2004 slowed business operations to a trickle [1].

The new system, designed to improve financial management, did not work out as planned. The *Federal Times* quoted one Federal Technology Service (FTS) procurement worker as saying, "People are so upset that they can't figure out how to do their jobs on the new system that someone bursts into tears almost hourly." FTS commissioner Sandra Bates was aware of the frustration: "I know it is very, very difficult to learn the new system... It is one of our lessons learned. We gave everyone extensive training before the system went live but didn't realize employees would need the trainers there while they started using the new system."

The system is unnecessarily complicated to use. Instead of being able to save a file with a few clicks, employees now must learn 15 steps. Bates was also quoted in the *Times* as saying, "The system is not simple, and how to do things is not always intuitive. The problems people are having are not trivial, but we have to work through them by getting [people] more training."

The UK Passport Office experienced similar problems in 1999 when installing a new system for issuing passports. After installation, a backlog of passports started building up, eventually leading to delays of up to three months to obtain a passport. Reasons for the loss in productivity included the need to correct errors in scanned data and the large number of keystrokes and onscreen operations required. The test program had not extended to thorough testing of the system's impact on productivity.

WHAT WENT WRONG? In both cases, the main problem was the lack of consideration for usability during design. Both were complex systems, and the impact of usability on productivity required consideration. Neither development process gave sufficient attention to designing for the user.

In its report [2] on the passport delays, the UK National Audit Office recommendations included this remark:

Organizations should pay special attention to the interaction between the new system and those expected to use it, and take into account users' views on the practicability and usability of the new system.

How could and should usability measures have been introduced into the development process to address these problems?

HOW TO FIX IT: INTRODUCING MEASURES. The newly published ISO/IEC 25062 Common Industry Format (CIF) [3] provides tools to introduce usability measures into development. The CIF format is designed for reporting results of formal usability tests

with quantitative measurements and is appropriate for comparative testing.

By applying the CIF methodology at three key points, the GSA and UK passport offices would not have been surprised by the poor usability of the delivered systems:

1. Measuring the usability of the system currently in place, thus identifying a baseline
2. Identifying and specifying target usability requirements from the baseline and users
3. Measuring the usability of the new system

Although these three steps may appear intimidating, adopting the CIF and quantitative measurement is straightforward and easy.

STEP 1. MEASURING THE USABILITY OF THE CURRENT SYSTEM

Which measures matter? The CIF adopted the ISO 9241-11 [4] definition of usability: “Usability is the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use. Thus, the CIF employs effectiveness, efficiency, and user satisfaction as three areas of measurement.”

- Effectiveness is a measure of the accuracy and completeness with which users achieve specified goals. Common metrics include completion rate and number of errors.
- Efficiency is a measure of the resources expended in relation to the accuracy and completeness with which users achieve goals. Efficiency is related to productivity and is generally measured as task time.
- Satisfaction is the degree to which the product meets the users’ expectations—a subjective response in terms of ease of use, satisfaction, and usefulness.

Although the CIF suggests measures for each of these metrics, it is important to note that the CIF does not detail what to test. Instead, the CIF codifies best practices for describing and reporting on the usability test so the CIF is flexible and adaptable for the practitioner.

How is usability measured? In both examples, new systems were replacing existing systems. This facilitates creating a usability baseline. To establish a baseline, measure usability with a summative test. It is essential to identify three components first:

- Primary users: Who are the main users? What are their characteristics?
- Primary tasks: What are the main or representative tasks?
- Context of use: What is the user’s computing environment (computer configuration, OS, settings, browser...)?

It is essential to identify and recruit participants representative of the system’s users, who complete key tasks in an environment that simulates real working conditions as closely as possible. Take measures of efficiency, effectiveness, and user satisfaction. Document results using the CIF format.

Identify the primary users, primary tasks and context of use. What is the users’ background: What are their goals, job responsibilities, daily activities/tasks, workflow, time constraints, and schedules? What is their computing environment; what other applications do they use routinely? What do they like and dislike about the current applications?

Let’s revisit the GSA system. In this environment, the primary users are procurement officials, buyers, accountants, pay personnel, and approving officials under tremendous pressure to process orders and payments as quickly as possible. Most are in cubicles and have the standard US-government desktop configuration.

Identify measures. Based on user interviews, identify appropriate measures.

Effectiveness: The CIF suggests percent task completion, error frequency, frequency of assists to participant by testers, and frequency of accesses to help or documentation during the tasks. Again, the CIF provides flexibility to the usability practitioner to define the



REFERENCES 1. *Federal Times*. 2004. New computer system brings work to a crawl at 2 GSA regions. July 21, 2004. 2. *National Audit Office*, 1999. United Kingdom Passport Agency: The passport delays of Summer 1999. Publication HC 812 1998-1999. 3. ISO/IEC 25062 2006. *Software Engineering - Software product Quality Requirements and Evaluation (SQuaRE)-Common Industry Format (CIF) for Usability Test Reports*. 4. ISO 9241-11. 1998. *Ergonomic requirements for office work with visual display terminals (VDT)s - Part 11 Guidance on usability* (ISO). 5. IUSR 2006. www.nist.gov/iusr (accessed July 2006)

level of the task and the success criteria for task completion.

Taking our GSA example, is the task “entering a correct order for a piece of equipment,” or “saving the file for ordering the equipment?” In either case, how you define complete and the errors are important. Assess task completion and errors against objectively defined criteria.

Efficiency: A common metric of efficiency is time on task, appropriate to our GSA example. Other measures of efficiency can also provide insight.

Consider the 15 steps required to save the GSA file. Or the number of pages or screens accessed, the number of files required, the number of fields that require data input—but the metric should relate back to the users and their goals.

Satisfaction: Measure satisfaction with a questionnaire. The CIF identifies a number of widely available questionnaires. Or develop an internal instrument. In any case, the questionnaire should consider including measurements of satisfaction, ease of use, and usefulness.

Establish a baseline. After gathering all this information, you have identified the key components required to perform a usability test. Document the results of the usability test using the CIF.

These baseline values can influence target values or goals for the new system. However, it’s not enough to just plug these baseline values into the requirements for the new system. The usability practitioner must investigate if these baseline values are representative. Do they meet the users’ goals and expectations? Are users’ expectations realistic, too high, or too low? To establish appropriate targets for the new system, assess the adequacy of the baseline measures.

STEP 2. SPECIFYING NEW USABILITY REQUIREMENTS. Use baselines for the existing system to establish usability requirements for the new system. To support this process, the same group sponsored by NIST that produced the original CIF recently developed a Common Industry Specification for Usability-Requirements (CISU-R) [5].

The CISU-R provides a structure for:

- Defining usability requirements in sufficient detail to make an effective contribution to design and development
- Defining usability criteria that can be empirically validated subsequently and reported in the CIF format

Using this approach, you can specify requirements for usability and assess the findings of a usability test of the new system against the original requirements.

The CISU-R is intended to support proactive collaboration between a supplier and customer to identify how a product can be effective, efficient, and satisfying. The CISU-R has three parts:

1. *The context of use:* intended users, their goals and tasks, associated equipment, the physical and social environment in which the product will be used, and examples of scenarios of use.

In the GSA example, training was a significant obstacle. In the context-of-use section, the amount of training thought necessary to complete the users’ tasks would be specified. For example, the amount of training needed for the new system should be no more than what the former system needed.

2. *Usability measures:* effectiveness, efficiency, and satisfaction metrics for the main scenarios of use with target values where feasible.

Users of the new GSA system complained that saving a file took 15 steps instead of the couple of steps in the original system. Baselining the original system could inform an efficiency requirement that the new system could not require more steps than the original system.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without the fee, provided that copies are not made or distributed for profit or commercial advantage, and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on services or to redistribute to lists, requires prior specific permission and/or a fee. © ACM 1072-5220/06/1100 \$5.00

3. *The test method*: the procedure to test whether the usability requirements have been met, and the context in which the measurements will be made. This provides a basis for testing and verification.

The test method for the usability requirements should reflect the context of use and the usability measures. The test environment and tasks should be as close to the operational environment as possible to help ensure the validity of the test results. If the GSA system user originally took three steps to save a file, saving a file should be one of the test tasks. The number of steps that most users took with the new system can then be compared with a requirement of no more than three steps.

Each of the three parts of the CISU-R can be part of an aggregate sequence. Context of use can be used without usability measure and test method if no measurable values can be established. Context of use and usability measures can be specified even when no formal testing is planned.

STEP 3. MEASURING THE USABILITY OF THE NEW SYSTEM. It is important to replicate the usability test of the old system on the newly developed system and document the results in a CIF report. After completion of the CIF, the baseline metrics and the target measures from the usability requirements specified in step 2 can be compared. Now you can address the question, "Is the new system usable?"

CONCLUSIONS. Using the CIF to specify usability measures can positively influence the development process. The CIF remains inherently flexible and adaptable—empowering the practitioner to identify success criteria specific to the application.

Summative usability testing with objective user performance and subjective satisfaction metrics based on existing systems provide an effective way to communicate usability requirements and usability assessments. The CIF standardizes the information captured about usability testing with users. The newly introduced CISU-R provides a structure for specifying usability requirements. These specifications provide the tools to design the measures and collect the metrics for determining the usability of systems.



ABOUT THE AUTHORS *Mary Frances Theofanos is a computer scientist in the Visualization and Usability Group at the National Institute of Standards and Technology where she works on the Industry Usability Reporting Project developing standards for usability and studies the usability of biometric systems. Previously, she was the Manager of the National Cancer Institute's Communication Technologies Research Center, a state-of-the-art usability testing facility including an extensive research program on the intersection of accessibility and usability. Mary spent 15 years as a program manager at the DOE's Oak Ridge National Laboratory complex.*



Brian Stanton is a cognitive scientist in the Visualization and Usability Group at the National Institute of Standards and Technology where he works on the Industry Usability Reporting Project developing usability standards and investigates biometric and robotic usability. Previously, he worked in private industry designing user interfaces for B2B Web applications.



Nigel Bevan is a usability consultant based in London and a research fellow at the University of York. He has managed several European projects to develop and apply usability methods and metrics. Bevan was a major contributor to the CIF and editor of the CISU-R, and has worked on many international standards. He is director of professional development at the UPA and leads the UPA Usability Body of Knowledge project.