

Comparative Evaluation of Usability Tests

Rolf Molich, DialogDesign (Denmark), molich@acm.org (Moderator)
Nigel Bevan & Ian Curson, National Physical Laboratory (UK), usability@npl.co.uk
Scott Butler, Rockwell Software (USA), sabutler@software.rockwell.com
**Erika Kindlund & Dana Miller, Sun Microsystems, JavaSoft Division (USA),
erika.kindlund@eng.sun.com**
Jurek Kirakowski, HFRG, University College Cork (Ireland), jzk@ucc.ie

1. ABSTRACT

Four commercial usability labs have carried out a professional usability test of the same calendar program. This paper discusses similarities and differences in process, reporting and results.

2. INTRODUCTION

There are two basic forms of comparative usability testing:

1. One professional lab tests several applications within the same subject area
2. Several professional labs test the same application.

Form (1) is rather well-known. However, there are few (if any) documented comparisons of professional usability tests.

Usability testers constantly criticize the work of others - and serve a useful purpose in doing so. As professional usability testers we must then be very careful to ensure the accuracy, usefulness and usability of our products. Swallow our own medicine, so to speak.

This paper presents the results of a comparative evaluation, where four commercial usability labs have carried out a professional usability test of the same commercial calendar program. The purpose of the comparative evaluation was to observe the different ways in which four independent laboratories approached the task of carrying out a cheap usability test to test the usability of a piece of software for new users.

The purpose of the comparative evaluation was not to pick a winner. It would be unfair even to attempt to do so. Each approach and report has its strengths and weaknesses. The purpose of the exercise was to

- Demonstrate the variety of some of the many different approaches to professional usability testing that are being applied commercially today.
- Show each team the comparative strengths and weaknesses of its approach to conducting and reporting usability tests.
- Provide an informed basis for an ongoing discussion of whether professional usability testing is an art or a mature discipline that turns out reproducible results.

This paper compares process, reporting and results. The paper discusses the difference between usability testing and good usability testing.

3. WHAT WE HAVE DONE

The following usability labs participated in the evaluation study (in alphabetical order):

- HFRG (Human Factors Research Group), University College Cork (Ireland)
- National Physical Laboratory (UK)
- Rockwell Software (USA)
- Sun Microsystems, JavaSoft Division (USA)

The usability labs that participated in the evaluation study reacted on a posting from the moderator on the UTEST listserv.

The program used in the test was the English language version of Task Timer for Windows, version 2 (TTW). TTW is a calendar program written by the Danish software house DSI for the Danish company Time/system. Time/system

manufactures traditional paper calendars but has seen an increasing market in computerized calendars. TTW exists in a number of different language versions.

To avoid licensing and copyright problems a demonstration version of TTW was used for the test. The only difference between the full version of TTW and the demonstration version is that the demonstration version can be started only 50 times.

TTW was selected for the test by the moderator. TTW was not known by any of the labs before the test. The software was made available to the participating labs 3-4 weeks before the reports were to be delivered to the moderator.

To avoid any negative reactions from the manufacturer of TTW the moderator took the following precautions:

1. Obtained the permission of the TTW product manager to conduct the test
2. Used an outdated version of TTW – many of the problems found in the test have been corrected in later releases.
3. Used a demo version of the program that is given away to any one who asks for it and is distributed freely at trade shows.

The usability test scenario was:

Time/system® is a Danish company that manufactures and distributes paper calendars. In the fall of 1994 Time/system released Task Timer for Windows version 2 as a computer version of the paper calendar.

The primary user group for TTW is professional office workers, typically lower and middle level managers and their secretaries. Time/system also offers the demo version of TTW freely to anyone at hardware and software exhibitions, conferences, and "events", e.g. Microsoft presentations. Time/system hopes that the demo version will catch the interest of people who pick it up by chance.

TTW is intended for users who have a basic knowledge of Windows. Familiarity with the paper version of the calendar or with other electronic calendars is not required.

Time/system is planning to send out version 3 of TTW in April 1998. However, their sales staff have heard negative comments about users' initial experience with the program, and TTW faces stiff competition from other programs, like Microsoft Schedule.

They have therefore asked you to perform a cheap usability test involving e.g. five typical users to test the usability of the software for new users of the product.

Task Timer for Windows is a diary, task and project management program for individuals and work groups. To reduce cost, you have agreed with Time/system to focus mainly on the diary and address book functions for individuals. In other words: Do not test task management, project management, networking functions, etc.

Each usability lab was asked to carry out a "normal" usability test of TTW and report the results in a usability report. Each lab was asked to use its standard usability report format with one exception: The name of the company should not be directly or indirectly apparent from the report. Therefore, the usability labs are referred to as Team A, B, C, and D in the following.

In addition, each usability lab was asked to report in an addendum

- Deviations from its standard usability test procedure.
- Resources used for the test (person hours).
- Comments on how realistic the exercise had been.

The labs did not communicate during the test period.

After all tests had been completed and the test reports had been received by the moderator, copies of all reports were distributed to each of the participating labs for inspection and commenting. Collecting the reports turned out to be a quite laborious task. Only team C delivered exactly as promised. Team B delivered one week later than promised. Team A and D revised their schedules extensively during the test.

The moderator then prepared a first draft of this paper. The draft was circulated until there was reasonable agreement on the contents. In places where full agreement could not be obtained, we have included alternate opinions.

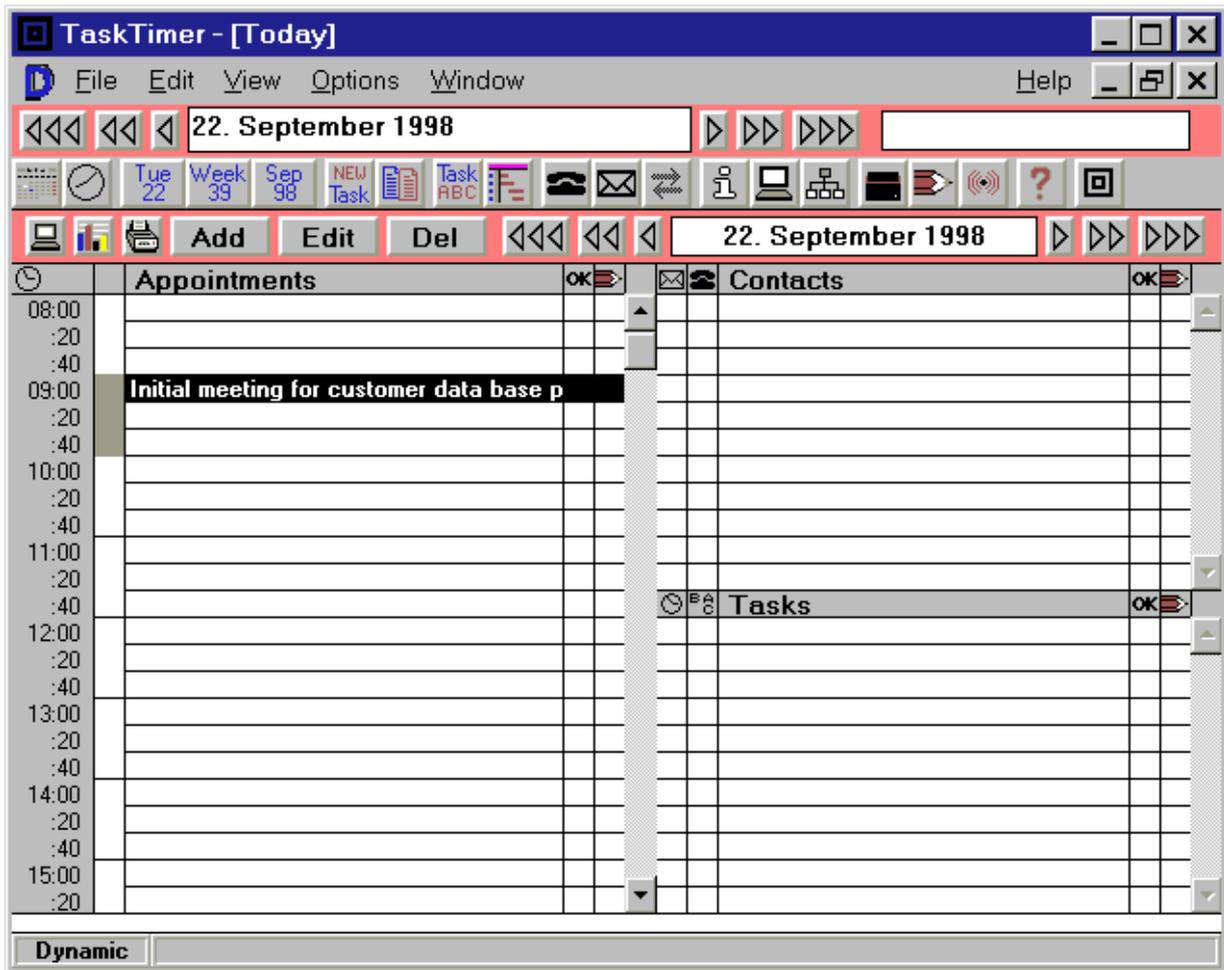


Figure 1. Overview of calendar for one day. Note the icons in the upper part of the screen, which were difficult to understand for many users.

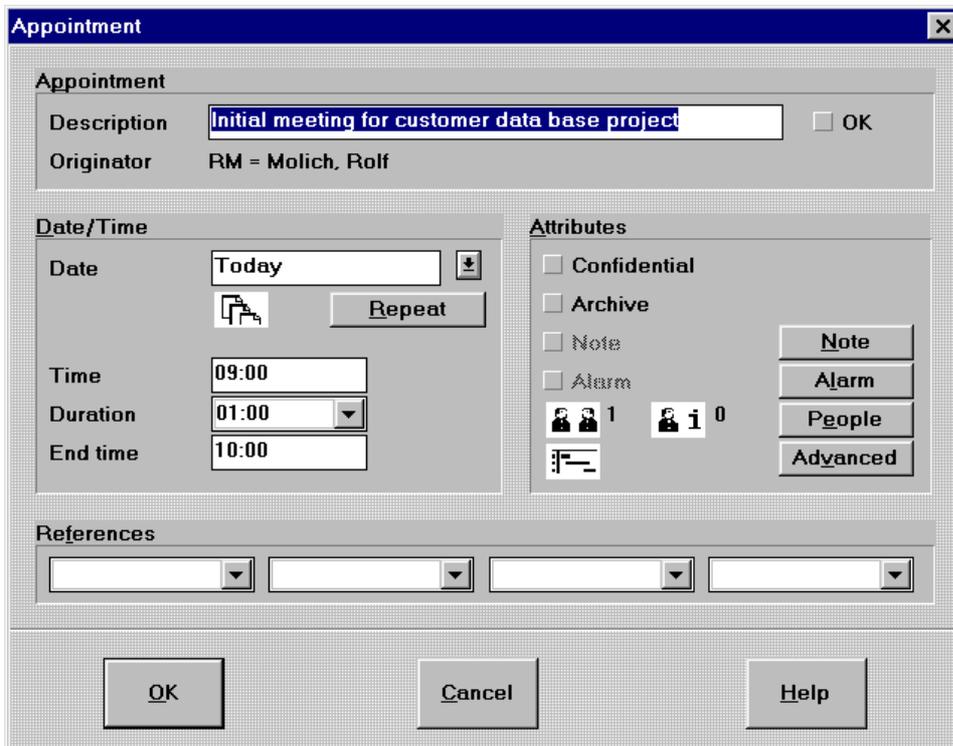


Figure 2. Dialog box for entering information about an appointment (the appointment shown in figure 1).

4. ABOUT TASK TIMER FOR WINDOWS

The manufacturer (Time/system) describes TTW in the following way:

TaskTimer for Windows is a PC application that is revolutionising office work. It's a diary, task and project management program, perfectly suitable for both individuals and groups.

Task Timer uses the PC to structure and interrelate all your information so that you get a clear overview of what has to be done, and who is involved.

Designed for the Network: TaskTimer for Windows is designed as a network program. It smoothly links people and gives them an efficient communications channel that keeps them all updated. Everyone on the network has complete access to all information, unless it is defined as confidential.

Printout: TaskTimer for Windows is designed to be completely compatible with paperbased planning systems, and anything that you see on the screen can be printed on paper in various formats.

The screen shots from TTW in figure 1 and 2 are included to provide an idea of the look-and-feel of the TTW interface.

5. INTERPRETATIONS OF THE SCENARIO

Team A interpreted the instructions in the usability test scenario "to focus mainly on the diary and address book functions for individuals" such that only these functions were to be tested, i.e. testing of other basic functions like installation, login etc. was given low priority. Team A also interpreted the instruction phrase "They have therefore asked you to perform a cheap usability test" to mean that it was the sales staff who had asked for the usability test, and neither management nor the developers.

The manufacturer's use of the phrase "diary" caused some difficulties for team C. In accordance with American English use of this word, the team assumed that the word meant "a book for keeping daily records of the writer's own experiences or observations" instead of "a calendar" as intended by the moderator and the manufacturer. When team C delivered its report to the moderator this use of the word was spotted and the other teams were warned about the potential misunderstanding. As a result some of the problems identified by team C are unrealistic, since the team attempted to use TTW for a task for which it is not well suited.

6. THE TEST PROCESSES

Table 1 presents comparative data about the usability test processes applied by the teams.

Team	A	B	C	D
1. Total person hours used for the test by the usability professionals. Test participants' time is not included. Equal to the sum of the following rows 2-4.	26	70	24	84
2. Time used for planning and usability context analysis	9	10	6	28
3. Time used for recruiting test participants and testing. Test participants' time is not included.	12	20	8	21
4. Time used for analysis of results and reporting	5	40	10	35
5. Number of usability professionals involved	2	2	1	3
6. Number of tests	18	5	4	5
7. Approximate length of each usability test in minutes	4 to 32	120	120	60
8. Profiles of test participants reported	No	No	Yes	Yes
9. Number of scenarios/tasks used in test	5	11	5	4
10. Detailed scenario descriptions provided (see also table 5)	Yes	Yes	Partly	Yes
11. Quantitative assessment of user interface provided	Yes	No	No	Yes
12. Results of heuristic evaluation performed by usability professional included in report	No	No	Yes	No

Table 1. Comparison of usability test processes.

6.1. Quantitative usability measurements

The purpose of a usability test report can be to enable management to make informed decisions about whether a piece of software should be released or revised from a usability point of view.

Team A and D provided quantitative assessments of the usability of TTW, which are useful for this purpose. Both measurements are based on the SUMI questionnaire and can thus be compared – see figure 3.

The SUMI questionnaire provides numeric assessments on the following scales:

- Efficiency: degree to which user feels he gets his work done well with the system
- Affect: degree to which user feels the system is enjoyable and stress-free to use
- Helpfulness: degree to which user feels the system helps him along
- Control: degree to which user feels in control of the system, rather than vice versa
- Learnability: degree to which user feels he can learn new operations with the system

There is also a Global usability score, which is a combination of items from each of the above scales.

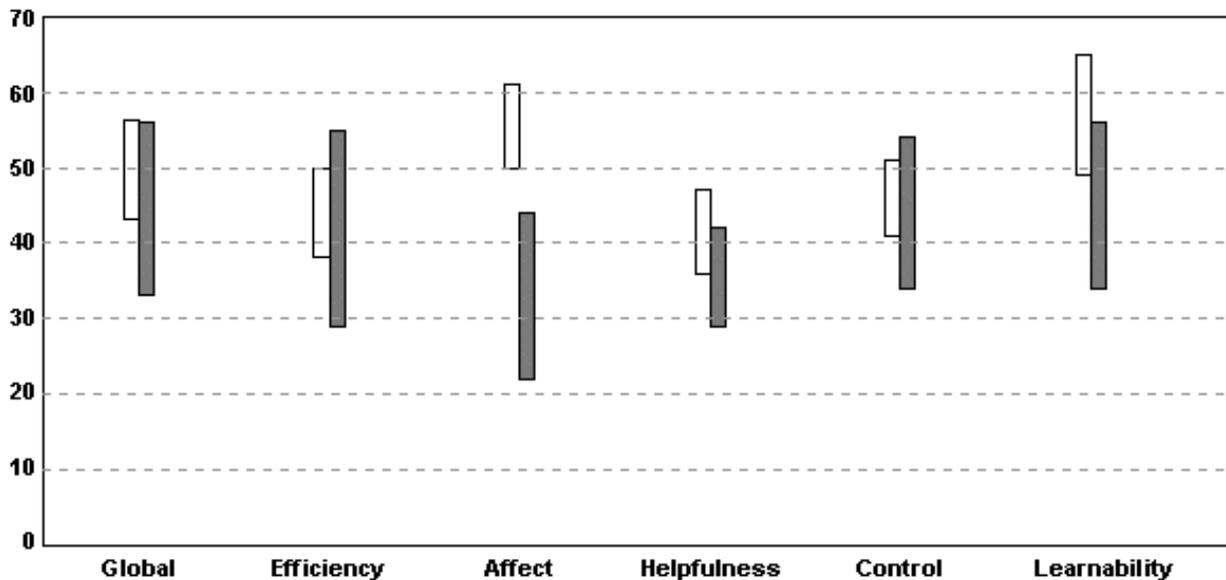


Figure 3. Comparison of quantitative usability measurements. Each column represents the 95% confidence interval around the median (the median is not shown) – that is, the range within which we are 95% certain that the true median of the user population lies.

White columns represent results from Team A, grey columns represent results from Team D

It should be noted that Team D used far fewer usability test participants than did Team A, which probably explains the larger 95% confidence intervals shown in Figure 3. Comparison between Team A and Team D results using SUMI must also take account that Team D included the software installation task and Team A did not. Since both teams reported the number of users involved and the tasks evaluated it is possible to make a meaningful interpretation of the differences in the two profiles.

6.2. Qualitative reporting

The purpose of a usability test report can be to enable developers to improve the user interface.

Team B, C and D provided qualitative problem reports, that is, descriptions of specific interface designs that caused problems for the users. Problem reports are useful for developers who want to improve the user interface.

6.3 Observations

1. The Usability Test Scenario said: "The primary user group for TTW is professional office workers, typically lower and middle level managers and their secretaries. TTW is intended for users who have a basic knowledge of Windows. Familiarity with the paper version of the calendar or with other electronic calendars is not required."

Team A, B and D appeared to select users on this basis. Team C selected users of "online and hardcopy contact management/calendar tools".

2. The methodology used by teams B, C and D was broadly similar: they all selected 4 or 5 users, set them realistic tasks, then observed the users performing the tasks, noting and interpreting any difficulties encountered.

Team A took a different approach, recruiting 19 users in 2 groups (distinguished by the extent of their experience of Windows 95). The users were also set realistic tasks, but the results were not obtained from observation of the users, but instead from the answers given to various questionnaires. Among other things the users were asked what they thought was the best, most favorable aspect of the software, and what they thought was the worst or least favorable aspect. This resulted in less feedback than the other approaches.

3. The SUMI results from teams A and D are broadly similar. This is interesting since the results were obtained by completely independent teams.
4. The only exception are the results for "Affect", which led to different conclusions. Team A reported: "The strongest element of Task Timer was that the users quite liked it: this score came in above the market average", while Team D found: "The results of the SUMI questionnaire, completed by the user after installation and familiarisation, are poor. In particular, the users did not like the software".

Team A comments that this difference provides a straightforward message to the developers: "The installation task really puts users off the software. This conclusion can be made with relative certainty because teams A and D explained their processes in detail, and this is one of the big differences between [their processes]. This could be a major disincentive to achieving sales to people who "people who pick it up by chance". "

However, the qualitative results from team C and D, who both tested the installation scenario, do not suggest specific usability problems of a severity that would justify such a conclusion.

5. There were wide ranges in the time reported by the teams: from 26 hours by Team A and 24 hours by Team C to 70 hours by Team B and 84 hours by Team D. Both team A and team C showed remarkable productivity.

Team C managed to produce an impressive 25 page report including log in only 10 hours.

Team A took 26 hours to test 18 users. It apparently took 40 minutes per user to administer a 20 min task and 3 questionnaires which took 10 min to complete. It then took an average of 3 minutes to analyse each questionnaire, and two hours to produce a 38 page report.

Team A comments that the use of standardised tools and software is the key to their increased efficiency. However, they also add that this increased efficiency comes at a cost: less attention can be given to specific diagnostics of poor interface features.

7. THE TEST REPORTS

Table 2 presents comparative data about the usability reports. The table is based in part on the recommendations provided in [1] and [2].

Team	A	B	C	D
1. Number of pages in report, excluding blank pages	38	18	25	66
2. Number of pages in main report, excluding blank pages and appendices	8	18	12	22
3. Number reports submitted	1	1	1	2
4. Length of executive summary in pages. An executive summary is recommended in [1] and [2]	½	Not provided	2 (Entitled "Human Interface Targets")	1½
5. Number of screen shots provided to illustrate problems	0	0	11	0
6. Quotations from test participants provided Recommended in [1]	Yes, in appendix	A few	Included in detailed 8-page log	No

Table 2. Comparison of usability test reports.

7.1 Observations

1. A usability test report that is ignored by the developers is useless. Therefore the usability of the usability report is important.
2. The report must be short. It appears that all teams agree that the main body of the report must be 8-22 pages.

The reports from team A and D contain a lot of detailed information about the SUMI method. This information has wisely been put into appendices or into a separate report. Nevertheless, the total size of these reports is considerable, and there is a risk that developers will not even look at it sufficiently closely to realize that they do not have to read it all.

Team A comments: "By putting this information in an appendix, we are quite clearly saying to our readers: "You don't have to read this unless you want to." We cannot help readers who don't even open a report... yet you will notice for the Team A report the first thing you see as you open the front page is "Summary of Findings and Recommendations" "

3. The report must be easy to understand. All reports lived up to this requirement with one exception. The reports that used the SUMI method contained some statistical information that can be hard to understand. Also, the SUMI method compares the usability of TTW to a large body of programs whose usability serve as a standard reference. This process is complicated, and may be difficult to sell to sceptical developers.

Team D disagrees: "Actually, we find quite the contrary is true - developers usually love SUMI results and can't wait to get hold of them! A score of above or below 50 is easy to interpret as a simple pass/fail criterion for usability! We usually have to invest quite a bit of effort in dissuading developers and managers from taking SUMI results at face value and jumping to conclusions."

Team A also disagrees: "The quantitative data that Team D and Team A have shown in the body of their report is the absolute minimum necessary to allow a reader to understand what is going on. There is a graph, and there are verbal conclusions. Everyone can see from the graph that most of the profiles are 'below the line' (i.e. 50). I think you are making too many negative assumptions about the consumers of this report. In my experience, sales staff are extremely receptive to these kinds of statistics, and they are well used to market survey results."

4. A usability report should have a professional and attractive layout. While all four reports have a professional layout, the report produced by team C is particularly attractive. See the excerpt in figure 4. Because of the generous use of screen shots, the report by team C is the only report that is reasonably comprehensible to a person, who does not know TTW. As argued in [1] this makes the report more useful for managers, other product teams, other usability teams, and perhaps even for the developers if they re-read the report after some time.

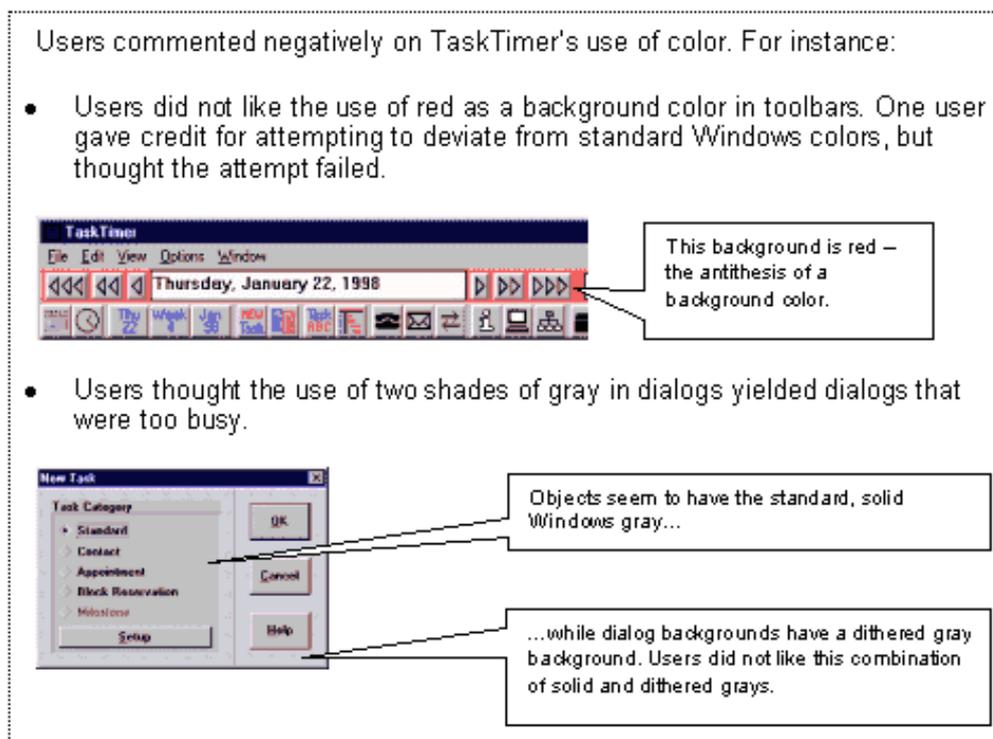


Figure 4. Excerpt from the usability test report produced by team C. Note the use of graphics to make the report more attractive and interesting. The text is comprehensible without access to the software.

5. TTW contains a large number of usability problems. Part of the reason for this may be that Time/system has never invested in usability, and usability was largely unknown in Denmark in 1993-1994 when the tested version of TTW was produced. It appears that team B got a little irritated about the general level of usability of the program. A quote: Default window size does not display Add/Edit/Delete buttons!!!! These functions are not supported in the menus and direct manipulation for delete is not apparent - users tried "backspace" key, not the "delete" key, on an expanded keyboard.
Although this problem is indeed serious, there is no need to use !!!! . Such exclamation marks are used several times in problem descriptions in the report from team B. Instead, it might be reasonable to draw management's attention to measures that could have prevented many of the usability problems found.
6. None of the reports included timing or task completion percentages.

8. THE TEST RESULTS

Table 3 presents comparative data about the usability test results.

Team	A	B	C	D
1. Number of reported problems	4	98	25	35
2. Number of reported problems that include specific recommendations for improving the interface	0	24	6	35
3. Number of reported problems that were encountered by one user only	0	4	2	8
4. Number of reported problems that deal exclusively with aesthetics (choice of colors, etc.)	0	0	5	1
5. Problems classified by severity Recommended in [1]	All four problems are severe	No	No	No
6. Number of positive findings reported. Recommended in [1]	1	4	3	0
7. Number of reported suggestions from test participants for improving the interface	0	2	5	0
8. Number of program errors reported	0	1	0	0
9. Indication of how many users encountered each problem Recommended in [1]	Yes	No	No	No

Table 3. Comparison of usability test results.

	Team A	Team B	Team C	Team D
Team A	-	2	1	1
Team B	2	-	3	8
Team C	1	3	-	5
Team D	1	8	5	-

Table 4. Problems found by more than one team. The table shows for instance that eight problems were found by both team B and D. Only one problem was found by all four teams. Another problem was found by three teams, namely B, C and D. Eleven problems were found by two teams.

Comparative Evaluation of Usability Tests

	Installation	Log-in	Familiarization with TTW	Basic Calendar	Advanced Calendar (Recurring appointment; group appointment)	Basic Address and Telephone Book
Team A				✓		✓
Team B				✓	✓	✓
Team C	✓	✓	✓			✓
Team D	✓	✓	✓	✓		✓

Table 5. Scenarios used by the teams. In the familiarization scenario the user is asked to take a few minutes to explore TTW.

8.1 Observations

1. The overlap between problems is remarkably small. The teams discovered a total of 162 problems (table 3, row 1). Thirteen problems were found by more than one team as shown in table 4. This means that 141 different problems were found (the figure 141 is difficult to verify based on the information given in the above tables, among other things because in some cases one problem found by one team matched several problems found by another team).

Even though team B found 98 problems and team D found 35 problems, only eight of these problems overlap, i.e. were found by both teams. Part of the explanation is that the teams did not use the same scenarios. As shown in table 5, team B for instance did not test the installation procedure and consequently had no chance of finding the problems associated with the installation procedure.
2. The reports include only few positive findings. A usability report should include positive findings in a reasonable proportion to the problems. Positive findings are important both to insure that developers do not remove features that the users liked, and also to increase the acceptance of the usability report by developers.
3. If many usability problems are found, the usability professional should limit the number of reported problems by eliminating less important ones from the report. Although it may seem reasonable to report everything that was found, the occurrence of a large number of usability problems point to more serious problems that should be addressed in the conclusion. Simply listing all problems may cause the developers to reject the report completely. Although all findings by team B are correct the number of problems reported in their report may seem overwhelming and depressing to the developers.
4. Findings should be presented in a way that are logical and easy to understand without any usability background. "Opinions" should be avoided. Opinions could be problem reports based on findings by one user only (table 3, row 3), or on a heuristic evaluation performed by the usability professional (table 1, row 12). Problems found by one user only may be "false positives", i.e. features that do not cause problems for the majority of the users and therefore should not be corrected. Such problems should be considered very carefully noticing that while one user had the problem the remaining users did not experience it.

9. STRENGTHS AND WEAKNESSES OF EACH REPORT

- Team A
- + Large number of test participants
 - + Extensive quantitative analysis
 - + Some positive findings included
 - Each test participant worked with the system for only a short period of time.
 - Few error reports. It will be difficult for the developers to improve the software based on the report.
 - The method by which SUMI arrives at its results is more complicated to explain than the qualitative method
- Team B
- + Very thorough list of problems
 - + Problems are carefully described
 - + Good recommendations on how to solve the problems
 - Overwhelming list of problems. Only a few positive findings
 - Layout not very attractive. For instance no figures.
 - No summary or easily identifiable conclusion

- Team C
- + Attractive and professional layout
 - + Screen shots included as figures. Problem reports are comprehensible without access to the software.
 - + Some positive findings included
 - Authors personal opinions included (but clearly labelled as such)
 - "Cosmetic" problems are indistinguishable from "disasters"
 - No summary or easily identifiable conclusion
- Team D
- + Manageable list of problems
 - + Extensive quantitative analysis
 - + Clear presentation of problems
 - The total report is quite long
 - No positive findings reported
 - The method by which SUMI arrives at its results is more complicated to explain than the qualitative method

10. OBSERVATIONS ON HOW REALISTIC THE EXERCISE HAS BEEN

Team A: It is practice for us, when taking on work, to understand the requirements of the report we are to produce. This is usually done in a face-to-face meeting. Who is the report for, what will it be used for, what level should it be pitched at, and so on. When the report is produced, we usually do two drafts: a first for comment, and a second final version. The report presented here is the first for comment type version. It usually would take us about another four hours to get the necessary changes from the client and produce the final, polished version to the client's satisfaction. What we usually respond to after version one are requests for clarification, and when necessary, more analysis. After all, we extol the virtues of understanding user requirements and iterative design, and it makes sense to apply these to our own evaluation work.

We took the requirements of [the moderator] very literally since we had nothing else to go on. We assumed that the commissioners of the report were the sales staff ("sales staff... have therefore asked you...") and that what was needed was a "cheap usability test" to verify the "negative comments" that the staff were hearing. We interpreted the "users' initial experience" to mean the first 20 or so minutes of their experience with TTW, and that we were told "to reduce cost, you have agreed with Time/system to focus mainly on the diary and address book functions for individuals."

Team B: The process of this study completely conflicted with the methodology promoted within our organization. Typically, members across the product team are involved throughout the process of user evaluations. Their involvement in our studies (including such things as: providing information which feeds the study design, observing users, and participating in a post-study debriefing session) promotes an inter-disciplinary process which we strongly advocate. This process is not facilitated by isolated usability engineers evaluating products and delivering a stand alone report.

Lacking dialog with engineers, designers and marketing personnel, we felt that the TTW study was being conducted in a vacuum. We had a very narrow understanding of the scope of TTW in terms of implementation parameters, design goals, and target user requirements. This affected our ability to adequately focus our evaluation of the product.

We suggest the following improvements for future studies of this type:

Provide more background information on the product being evaluated. This should at least comprise detailed target user profile information, functional requirements specification, design goals and/or specifications, and a rough idea of any implementation constraints that should be considered to limit recommendations. Ideally, it would have been wonderful to have a specific - limited - list of user tasks to evaluate, which would have negated the need for the above items.

Consider finding a more compact product to evaluate. It was unrealistic to adequately evaluate this product because it was too large. Due to the powerful and extensive nature of TT, it was difficult to contain the scope of the tasks. Additionally, the fact that there were so many usability issues throughout the product made prioritizing the findings near impossible, considering the lack of product team involvement.

Team C In general, there was no real deviation from what we do as test administrators for a usability test. While we took total responsibility for writing scenarios, this wasn't a big deal because we normally have a great deal

of input into scenario development. And while we recruited users, this wasn't completely out of the ordinary because we've done this in the past to help out development teams who were overworked.

The big difference was the utter lack of involvement of the TaskTimer development team in the usability test process. This would not have been acceptable at our company and if we were contractors, we would not have agreed to this kind of relationship with a customer.

In our normal process, members of the development team are actively involved in scenario preparation, user identification and recruitment, and always attend usability evaluations. Usually, we have a programmer, a writer, and a marketing representative. As a result of lack of involvement on the part of the TaskTimer development team in this exercise, we cannot be sure if we've fully met their needs. For instance, we have no idea whether we've tested the journal function. If what we guessed to be the journal function [the diary, ed.] is not the journal function then we missed a major customer requirement for this test. We are also unsure whether our scenarios are robust enough.

Moderator: I basically agree with most of the above viewpoints. The most important purpose of a usability test is indeed to involve the development team in the study.

I maintain that TTW has a realistic size. Most of the software that I test for usability has a size where I can only cover a fraction of the total functionality within the resources allocated to usability testing. Our challenge as usability testers is to devise representative, orthogonal tasks that will illustrate a number of usability error types. We can then hope that the most serious problems will be corrected and that developers on their own will correct similar problems in the large parts of the software that we do not have resources to test. With this exercise, I wanted to see how the participating teams would handle this challenge – and I think that the teams handled it well. We can never get good or even acceptable usability just by testing.

Providing development team scenarios to the usability team is tempting and would of course ease the work of the usability team a lot. However, it would also increase the danger that important problems in the user analysis were not discovered because the scenarios were not questioned. An important part of usability testing is to provide an independent review of the task scenarios.

11. CONCLUSION

Three results of this comparative test have surprised us:

- The reproduceability of the SUMI results
- The large number of usability problems detected in the software
- The limited overlap between the usability problems detected in the software.

The limited overlap between the usability problems found by the teams may be a result of a large number of usability problems in TTW. It may also be a result of the different approaches to usability testing taken by the participating teams, in particular the selection of different usability test scenarios.

12. ACKNOWLEDGMENTS

The teams gratefully acknowledge the contributions to the evaluation by the following individuals:

- Richard Lakin and Andrew Harry from National Physical Laboratory.
- Elaine M. Gilman, Jennifer L. Giordano, Hans W. Kim and Myron M. Shawala III from Rockwell Software.

13. REFERENCES

- [1] Joseph S. Dumas and Janice C. Redish: *A Practical Guide to Usability Testing*; Ablex 1993
- [2] Jeffrey Rubin: *Handbook of Usability Testing*, John Wiley 1994

14. PANELISTS BIOGRAPHIES

Nigel Bevan is head of Usability Services at the National Physical Laboratory. NPL Usability Services offers a range of consultancy services to help improve the usability of products, and has provided usability tools, training and services to over 40 organisations world-wide. These includes use of methods for usability measurement developed in the European

MUSiC project. NPL coordinates a network of European Usability Support Centres set up with European Union funding.

Scott Butler manages the Human Interface group at Rockwell Software and has extensive experience with software, information, and hardware usability.

Ian Curson has been working at National Physical Laboratory since obtaining his MSc in Technical Communication and HCI in 1993. He has a background in software development and is responsible for usability consultancy, training and evaluation services at NPL.

Jurek Kirakowski graduated from Edinburgh University and since 1984 is the Director of the Human Factors Research Group at University College Cork, Ireland. As well as being closely involved with the setting up of the European Usability Service Centre network, Jurek Kirakowski is the originator of the Software Usability Measurement Inventory (SUMI) questionnaire as well as more recent questionnaires for multi-media and web site usability.

Erika Kindlund is a usability engineer at Sun Microsystems, JavaSoft Division. She evaluates emerging web technologies and their impact on the user experience. Erika Kindlund worked previously in Human Factors engineering at IBM and as a research scientist for the Interactive Multimedia Group at Cornell University. She has also edited three psychology documentary films aired on PBS.

Dana Miller is a usability engineer at Sun Microsystems, JavaSoft Division. She has a PhD in Psychology from Rice University.

Rolf Molich holds an M.Sc. in Software Engineering from the Technical University of Denmark from 1974. Rolf Molich has been working with practical software development in Danish industry ever since. He has been working full or part time with usability since 1984. Rolf Molich owns and manages DialogDesign, a small consultancy company specializing in usability. He has performed extensive usability testing, in particular of web-sites.