Motivating Change and Reducing Cost with the Discount Video Data Analysis Technique

Jody Wynn and Jeremiah D. Still

Missouri Western State University, Department of Psychology,
4525 Downs Dr., St. Joseph, MO 64507, USA
{jwynn,jstill2}@missouriwestern.edu

Abstract. Testing the usability of an interface is a critical phase of product development. However, it is often reported that analyzing the data from such testing consumes too many limited resources. We attempted to reduce this consumption by proposing a new technique, Discount Video Data Analysis (DVDA). We compared it with another popular accelerated analysis technique, Instant Data Analysis (IDA). Using IDA, evaluators analyze data after a series of usability tests, whereas DVDA calls for analyzing the data after every test in the series. Immediate analysis decreases the chance that subsequent test data will negatively interfere with evaluators' recall. Additionally, DVDA produces a video of the testing allowing the users' emotional responses (e.g., frustration) to be shared with developers who may be resistant to interface modifications. We found evaluators using DVDA identified more usability issues and provided more supportive evidence for each issue than evaluators using IDA.

Keywords: Data Analysis, Usability Evaluation, Discount Usability Testing.

1 Introduction

Two of the most prevalent usability evaluation methods used in the Human-Computer Interaction (HCI) community are expert reviews [1][2][3][4] and empirical tests [5][6][7][8]. The majority of studies comparing expert reviews and empirical tests indicate that the latter are superior for finding more severe interaction issues [9][10][11], but they are often discredited for being too time-consuming and expensive [4][9]. Ideally, HCI professionals need to have both the efficiency of expert reviews and the effectiveness of empirical tests. This need guided our development of the Discount Video Data Analysis (DVDA) technique.

The idea for DVDA was inspired by Kjeldskov, Skov and Stage's (2004) technique, which they coined Instant Data Analysis (IDA), of completing an empirical usability evaluation within a day [12]. The defining processes involved in IDA are to observe usability tests for a few hours then analyze and document usability issues later that day. Use of this procedure greatly reduces empirical testing costs compared with traditional methods.

Our DVDA technique diverges from IDA by stipulating that analyses occur after every test in a testing session rather than waiting until the end of the session. It has

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 321–328, 2011.

been shown that remembering new information can interfere with our ability to remember older information; especially when the information is highly similar [13]. Thus, when multiple tests are observed – as is done using IDA – subsequent tests could proactively interfere with the evaluators' recall of earlier usability issues. We predicted that documenting usability issues immediately following their occurrence ought to decrease memory errors.

Another difference between DVDA and IDA is that the former prescribes video recording the tests. These recordings may be employed to help resolve disagreements between study observers and to convey emotional aspects of the interaction – user frustration – to developers who might be resistant to changing an interface. Although the cost of video analysis may have seemed prohibitive in the past, the development of digital recording software makes it possible to record and edit videos in minutes rather than hours. To watch any portion of a video, the test evaluators can digitally "jump" to that point in the recording.

To examine if the differences between DVDA and IDA influenced the number of usability issues evaluators identified, both methods were used to analyze a commercial website. We predicted that evaluators using DVDA would identify more usability issues than evaluators using IDA. Additionally, we predicted that DVDA evaluators would produce more detailed reports without expending more resources than IDA evaluators.

2 Method

2.1 Participants

The university institutional review board approved all experimental procedures. Four undergraduate Introductory Psychology students were recruited to participate in a 20-minute usability test in exchange for course credit.

2.2 Facilitator and Evaluators

One Human Factors graduate student acted as the facilitator. Two Human Factors graduate students and two upper-level undergraduate Psychology students acted as usability test evaluators.

2.3 Apparatus and Stimuli

The usability tests were conducted in a private testing room equipped with a computer workstation. Participants interacted with a commercial news website. Remote observation took place in separate rooms, also equipped with computer workstations. Morae Usability Software [14] was used to record, remotely observe, and review the usability tests.

2.4 Procedure

To compare DVDA to IDA, usability test evaluators used both techniques to analyze a website (see fig. 1). The evaluators were paired - one graduate student and one

undergraduate student. Each pair employed either the IDA or DVDA technique. The DVDA evaluators followed instructions for the DVDA methodology (see Appendix A). The IDA evaluators followed instructions for the IDA methodology [12]. Both sets of evaluators followed our standard documentation for severity ratings (see Appendix B).

Both Techniques	**DVDA Technique**	**IDA Technique**
P: Participant	L: Data Logger	L: Data Logger
F: Facilitator	M: Monitor	M: Monitor

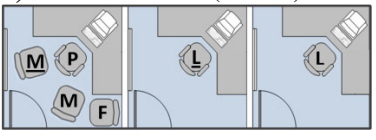
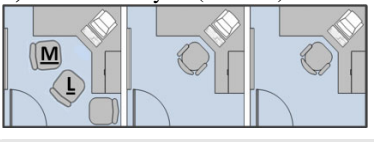
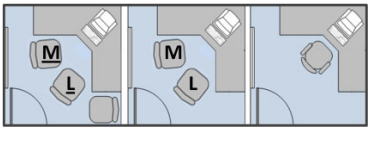
Layout	Outline
1) Both Collect Data (20 min.) 	- Facilitator read tasks - Monitors watched for errors - Data Loggers remotely watched for errors
2) DVDA Analysis (10 min.) 	- DVDA evaluators analyzed data - IDA evaluators did not meet or discuss the test
• Repeated steps (**1**) and (**2**) for the second, third, and fourth test • Monitors and Data Loggers switched roles before each test	
3) IDA Analysis (40 min.) Both Create Reports (20 min.) 	- IDA evaluators analyzed the data - DVDA evaluators created a report - IDA evaluators created a report

Fig. 1. The left column shows the room layouts with key actors. The right column outlines the experimental procedure used to compare the DVDA technique with the IDA technique. It is important to note that both accelerated usability techniques took the same amount of time. The DVDA evaluators spent 10 minutes after each of the four tests to perform their analysis; resulting in 40 minutes of analysis. IDA evaluators spent 40 minutes performing their analysis after all the tests were complete. They both took an additional 20 minutes to create their data analysis reports.

A facilitator acted as a liaison between the evaluators and the usability test participants. She instructed the participants to use the think-aloud usability test protocol [8], asking them to say what they are thinking, doing, and feeling while interacting with the website.

One test evaluator for each methodology sat in the room with the test participant and performed the role of test monitor. The monitors did not take notes, but the DVDA monitor wrote down the elapsed time of the recording when she recognized a usability issue. Sometimes she wrote a letter or word by the time notation to remind her of what the issue was (extensive notes were not taken).

The other two evaluators acted as data loggers and sat in separate rooms. They watched the test on their computer screens and took extensive notes. The DVDA data logger made a notation of the elapsed time of the recording each time he documented an issue.

During the 20 minutes allotted for each test, the facilitator instructed the tasks to be performed by the participant (e.g., search for grocery coupons, subscribe to a channel, place a furniture ad, etc.). At the conclusion of each test, the DVDA evaluators took ten minutes to analyze the test they had just observed (following the protocol outlined in Appendix A). During their analysis, the DVDA evaluators referred to the data logger's notes, the website, and the video of the usability test to aid their recall of the issues encountered by the participant. After ten minutes, the data loggers and test monitors switched roles and took their respective places in the testing room and observation rooms.

After the fourth and final 20-minute test, the DVDA test monitor and data logger spent 10 minutes on their last test analysis and 20 minutes creating a report. At this time, the IDA test monitor and data logger met for 60 minutes and analyzed the usability test and created their report.

All four evaluators finalized their data analyses, determined the severity of each issue, and typed their report within 60 minutes. The DVDA evaluators performed their 40-minute analysis in four 10-minute increments after each test, and then spent 20 minutes producing their report after the entire usability testing session was done. The IDA evaluators performed their 40-minute analysis after watching all four tests. Directly following their analysis they spent another 20 minutes creating their report. Once both teams completed their usability reports they discussed their error severity categorization. We hoped this would promote interobserver reliability across the two teams. Each team reviewed the other team's issue descriptions and agreed on the severity rating of issues they found in common.

The IDA report only included a description and severity rating of each issue. However, the DVDA report included a description, severity rating of each issue, the number of test participants who experienced each issue, and the time each issue occurred.

The IDA and DVDA techniques' performance were compared in terms of quantitative and qualitative measures. The quantitative data extracted from the reports were the total number of unique issues in each severity category, the number of issues in each category identified using DVDA and IDA, and the number of test participants that had experienced each problem. The qualitative measure was the thoroughness of the issue descriptions.

3 Results

Out of thirteen total issues, twelve were reported by the DVDA evaluators and ten by the IDA evaluators. The results for both methods were similar, the only differences

being in the "Serious" category. In that category, the DVDA evaluators discovered three issues that the IDA evaluators did not; conversely the IDA evaluators discovered one issue that the DVDA evaluators did not.

There was a qualitative difference in the content of the usability reports. The DVDA reports were more detailed than the IDA reports and included the number of users experiencing each problem. The DVDA reports also included a reference to the supporting videos via a timestamp on each issue.

4 Discussion

As expected, more usability issues were identified using the DVDA method than the IDA method. The key quantitative difference was that the DVDA method captured the number of test participants that experienced each issue. Information regarding frequency and severity of interaction issues is useful for prioritizing issues for redesign or reengineering [15]. Knowing which participants experienced each issue also allows the evaluators to identify outliers – those participants who show a trend for encountering more issues than other participants. Further, DVDA evaluators were able to include more descriptive elements in their report, and a video recording of each test was produced to support this information. The DVDA evaluators were able to explain the issues, indicate how many users had experienced the issue, and where the issue could be found in the video. Having the ability to reference the videos may become important if the evaluators need to justify the validity of the issues they report to developers.

These additional pieces of evidence are valuable for usability professionals, especially those attempting to gain credibility and respect from product developers. Nielsen [16] says, "As [a company's] usability approach matures, organizations typically progress through the same sequence of stages, from initial hostility to widespread reliance on user research." It is common for usability professionals to find themselves in a hostile environment where user-centered design and usability testing are viewed only as activities that create more work for developers. The mindset of the developers can be the greatest obstacle for deploying usability practices within an organization [17]. An effective way to convince developers of the importance of usability testing is to have them watch customers using their products [6]. When developers cannot attend the testing session, the video recording included in DVDA allows them to see users struggle with problematic design elements.

Although this study suggests that DVDA is a valuable technique there are some limitations. Due to time restrictions, only four 20-minute tests could be performed in this study. The evaluators only had time to find thirteen issues in total, but the DVDA technique revealed two more issues than the IDA technique.

We suggest a follow-up study in which the usability testing session is four or five hours long as prescribed by the IDA methodology. We would expect that given a longer delay between the experience and when it is retrieved from memory, significantly more usability errors will be forgotten by professionals using the IDA than the DVDA technique. Since the DVDA technique allows evaluators to record their insights when they are freshest in their memories, recalling the events of each test cannot be interfered with by observing other events in subsequent tests.

Both DVDA and IDA methods allow researchers to finish their analysis in a fraction of the time it would take using traditional methods. This study suggests that DVDA is capable of revealing more usability issues and producing a richer report than the established IDA technique.

Acknowledgments. We thank Sujoy Kumar Chowdhury, Joseph Grgic, Deanna Knox and Jacob Wharton for performing the usability evaluations.

References

1. John, B.E., Packer, H.: Learning and using the cognitive walkthrough method: A case study approach. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 429–436 (1995)
2. Frøkjær, E., Hornbæk, K.: Two psychology-based usability inspection techniques studied in a diary experiment. In: Proceedings of the Third Nordic Conference on Human-Computer Interaction, pp. 3–12 (2004)
3. Muller, M.J., Matheson, L., Page, C., Gallup, R.: Methods & tools: Participatory heuristic evaluation. Interactions 5(5), 13–18 (1998)
4. Nielsen, J.: Usability inspection methods. In: Conference Companion on Human Factors in Computing Systems, pp. 413–414 (1994)
5. Dumas, J.S., Redish, J.C.: A Practical Guide to Usability Testing (Rev Sub.). Intellect Ltd., Bristol (1999)
6. Dumas, J.S.: Stimulating change through usability testing. SIGCHI Bulletin 21(1), 37–44 (1989)
7. Frøkjær, E., Hornbæk, K.: Cooperative usability testing: Complementing usability tests with user-supported interpretation sessions. In: CHI 2005 Extended Abstracts on Human Factors in Computing Systems, pp. 1383–1386 (2005)
8. Rubin, J.: Handbook of usability testing: How to plan, design, and conduct effective tests. John Wiley & Sons, Inc., Chichester (1994)
9. Jeffries, R., Desurvire, H.: Usability testing vs. heuristic evaluation: Was there a contest? SIGCHI Bulletin 24(4), 39–41 (1992)
10. Karat, C., Campbell, R., Fiegel, T.: Comparison of empirical testing and walkthrough methods in user interface evaluation. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 397–404 (1992)
11. Thyvalikakath, T., Monaco, V., Thambuganipalle, H., Schleyer, T.: Comparative study of heuristic evaluation and usability testing methods. Studies in Health Technology and Informatics 143, 322–327 (2009)
12. Kjeldskov, J., Skov, M.B., Stage, J.: Instant data analysis: Conducting usability evaluations in a day. In: Proceedings of the Third Nordic Conference on Human-Computer Interaction, pp. 233–240 (2004)
13. Dewar, M.T., Cowan, N., Sala, S.D.: Forgetting due to retroactive interference: A fusion of Müller and Pilzecker's early insights into everyday forgetting and recent research on anterograde amnesia. Cortex 43(5), 616–634 (1900)
14. TechSmith Corporation, ©1995-2010, `http://www.techsmith.com/morae.asp`
15. Woolrych, A., Cockton, G.: Why and when five test users aren't enough. In: Proceedings of IHM-HCI 2001, vol. 2, pp. 105–108 (2001)
16. Nielsen, J.: Corporate usability maturity (2006), `http://www.useit.com/alertbox/maturity.html`

17. Bak, J.O., Nguyen, K., Risgaard, P., Stage, J.: Obstacles to usability evaluation in practice: A survey of software development organizations. In: Proceedings of the 5th Nordic Conference on Human-Computer Interaction: Building Bridges, pp. 23–32 (2008)

Appendix A: Discount Video Data Analysis (DVDA) Procedure

DVDA may be used with any empirical usability testing method which can be recorded via screen-capture software. The mechanics of the testing technique are not constrained by DVDA, although the following procedures must be included in the usability testing phase:

1. Video record the test using screen-capture software.
2. One evaluator act as test monitor and sit in the testing room with the test participant. Interact with the participant according to the testing method being used. Focus on the participant and do not take extensive notes. Notate the time each issue occurs, along with a letter or word to aid in recalling the issue later. If a subsequent participant encounters the same issue, put a tick mark beside the original notation to indicate replication of the problem.
3. One or more evaluators act as data logger(s) and take extensive notes, including the time each issue occurs. If a subsequent participant encounters the same issue, put a tick mark beside the original notation to indicate replication of the problem. Focus on documenting issues and do not interact with the test participant.
4. Evaluators must use a common notation time (e.g., recording elapsed time).

Data analysis is performed immediately after each test using the following procedure:

1. All evaluators meet at the conclusion of each test for approximately ½ the time the test took to collect the data. For example, if the test data collection took 60 minutes, spend no more than 30 minutes on the analysis.
2. Compare time notations, discuss experience, and document common issues.
3. For issues not found in common, explain the issue to the evaluator(s) who did not see it. If the evaluators agree it is a problem, document it.
4. To review any part of the test, "jump" (i.e., digitally skip) to that event in the video.
5. Count the number of tick marks by each issue and add one for the original issue to tally how many times each problem was experienced by a test participant. Include this information in the usability report.
6. Make screen shots if needed to augment the documentation.
7. Make video clips if needed to augment the documentation.
8. Categorize the issues based on severity criteria. See Appendix B for example severity criteria.
9. Perform a final review and compile a usability issue report that includes a description, the severity, and the number of participants who had each issue.

Appendix B: Usability Issue Severity Criteria

Critical: They could not reach the intended goal. This could be due to an error or failure in the website or because the path to complete the task was obscure.

Serious: They took an excessive amount of time to reach the intended goal. This could be because the path to complete the task was not obvious.

Aesthetic: They could complete the intended goal but were not pleased with an aspect of the look and/or interaction.