

Addendum

This Addendum reports on my evaluation criteria, my resources used for the evaluation, and comments on the evaluation

My evaluation criteria

I have used the following criteria for reporting incidents:

1. If a user commits an error (if it is a slip, it is not reported, but if the error is likely to have been caused by the interface design it is reported). Even errors that get corrected are reported.
2. If a user expresses worry, is surprised or suggest a different way to design the system
3. If the user clearly has a different interpretation of the system state than the actual state (e.g., "I have booked a truck", but no truck is booked)
4. If the user pauses for some time without knowing what to do
5. If the moderator helps the user on the way
6. If a task is not solved or not solved correctly
7. If the user says that s/he wants to call U-haul for help

My resources used for the evaluation

The table below illustrates the time I have used for the three categories asked for plus an additional category. I spent 17 hours in total being part of this study.

Preparation	100 minutes	Including reading instructions, orientate myself on the U-haul website, and preparing excel sheets
Watching videos	450 minutes	I watched the videos one by one, and I paused the videos whenever I made an incident note. I purposely did not attempt to create a unique problem list from the beginning but simply noted down incident by incident without matching an incident from one video with a similar incident from another (or the same) video.
Creating unique incident list	275 minutes	After creating five incident lists I started creating a unique incident list. While it has consumed additional man hours I feel it is the a more accurate process partly because I don't have to look for and remember whether I have seen a certain incident before while watching videos
Writing report	195 minutes	I wrote the report and the Addendum

Comments on the evaluation

Some thoughts that I had during the study

- The U-haul site is certainly mature and works very well. This may decrease the evaluator effect we see, simply because there are less complex problems, less bugs, and less ways users may get confused. That said, I'm surprised that I still identified many incidents
- The process of using a pre-defined excel sheet introduces a bias in the study. Firstly, some evaluators will be more accurate in this study than they usually are simply because the format is binding the user to report video foot stamps and other information that they may not include in their daily work. Secondly, the process may get evaluators to create a unique problem list from the start – that, I believe, is more complex but faster than making five separate incident lists, one per user session
- I do not do usability evaluations in my normal day-to-day job, and because I have been deeply involved in evaluator effect studies earlier on in my career, I may be slightly biased as a typical evaluator in this study
- I have no real experience in evaluating websites (I have evaluated mobile interfaces for many years)
- The style of the videos are great material for this type of study

Follow-up questions after submission of report

1. How familiar were you with the company U-Haul before we announced that we would use it for CUE-9?
 - a. I was familiar with the company. I have used it in one occasion 14 years ago.
2. How familiar were you with U-Haul's website before we announced that we would use it for CUE-9?
 - a. Not at all.
3. Approximately how many times have you rented a car?
 - a. Probably 15 times
4. Approximately how many times have you rented a car on the web?
 - a. Probably 8 times
5. How much time did you spend analyzing U-Haul.com before you watched the first video?
 - a. Around one hour
6. Approximately how many times did you pause the videos to deliberate or catch up with your notes?
 - a. Virtually every time I made an incident description
7. Did you watch all or parts of the videos several times? If yes, approximately how many times did you watch each part and how long were the parts of the videos that you watched several times?
 - a. For some incidents I re-visited the episode once or twice before I could understand what happened. After completing all five incident lists I realized that I some times had to go back to watching a video episode to create the unique incident list (i.e., collapsing several incidents into one unique incident description). On average I believe I have re-visited 15 episodes per user session. Each of these re-visits lasted approximately 15 seconds to 1 minute
8. Were there any burning questions that you would have asked the test participants during or after the video recorded sessions if you had been moderating the sessions? If yes, what were they?
 - a. The moderator acted professionally. Especially, she did not interrupt too early and she didn't let the user struggle for too long either. Overall, I did not have any burning questions. That said, it is clear that the moderator had a deeper insight into the system that I had and therefore asked users into areas I would not have done.