TECED user experience research and design

Multiple Facilitators in One Study: How to Establish Consistency

Laurie Kantner
TecEd, Inc.

Lori Anschuetz
TecEd, Inc.

ABSTRACT

In best practice for user research, a single researcher facilitates all study sessions to minimize variation. For larger studies, assigning one facilitator may miss an opportunity, such as catching select participants or delivering timely results. This presentation provides guidelines, with case study examples, for establishing consistency in multiple-facilitator studies.

INTRODUCTION

User experience research attempts to collect reliable data through realistic tasks with users from the target audience, using experimental methodology that minimizes variables affecting the data collected. One of these variables is the facilitator—how consistent the facilitator is from session to session, how neutral the facilitator remains in verbal and body language, and how detail-oriented the facilitator is in observing and taking notes. It's difficult even for a highly experienced researcher to collect data without biasing it.

No one in the field of user experience advises having multiple facilitators for a single study. Yet it occurs, as a reluctantly accepted compromise to the methodology. The most common scenario is a large-scale study requiring dozens of sessions, with stakeholders who need the results urgently and insufficient time for both planning and execution. Planning must not be short-changed; the remaining option is to shorten the calendar time allocated to execution (conducting sessions) by using more facilitators.

Another common scenario is a study in multiple locations where the team has research staff in those locations and does not want to budget for travel.

This presentation discusses how to reduce the risks that jeopardize the data in multiple-facilitator studies. It focuses on the methodology for establishing, not necessarily proving, inter-facilitator consistency by practitioners.

CASE STUDIES ILLUSTRATING TECHNIQUES IN THIS PRESENTATION

The following case studies illustrate the techniques for creating consistency among multiple facilitators.

Comparative Study of 2 Web Applications with 3 Facilitators and 60 Sessions

An Internet marketing company wanted to conduct a benchmark comparison study of its application for search engine marketing against a competitor, with statistical analysis of timing data, user difficulty ratings, observer difficulty ratings, mouse clicks, and errors categorized into 25 error types within 5 categories. Participants used one or the other application in 60-minute sessions. We determined that 60 sessions, 30 per application, were necessary to achieve the statistical confidence desired.

The client wanted the results in 6 weeks. Allowing time before sessions for planning (1 week) and recruiting and finalizing the script (2 more weeks), and time after sessions for data analysis and reporting (1.5 weeks), we were left with 1.5 weeks to conduct 60 sessions. At 5 sessions per day, we would need 12 session days (2.5 weeks) for 1 facilitator, assuming everything went perfectly.

We discussed the pros and cons with the client of assigning multiple facilitators to shorten the calendar time for sessions, and determined that we could control the disadvantages of multiple facilitators while achieving the advantage of meeting the deadline with high-quality results. It meant adding more time up front to ensure consistent facilitation.

Comparative Study of 2 Prototype Websites with 5 Facilitators and 400 Sessions

An Internet security company wanted to conduct a study eliciting user impressions about security of ecommerce transactions, comparing 2 prototype websites. In 1 prototype website, the ecommerce experience showed information about website validation, and in the other prototype website, this validation information was missing. Participants saw both websites in a 20-minute session, answering questions about their impressions along the way. Presentation of the websites was counterbalanced across the participants to mitigate order effects.

At the end, the facilitators asked participants to compare the websites and answer the key question of which website they were more comfortable doing business with. The client determined that 384 sessions were necessary to achieve the statistical confidence desired.

The client was firm about wanting the results in 6 or 7 weeks. We needed to spend 2-3 weeks up front vetting the test script to ensure an unbiased delivery of the information, and concurrently recruiting the participants. We needed to spend 2 weeks after completing the sessions to analyze and report the data, some of which was qualitative. That left 2-3 weeks to run the sessions. With 1 facilitator conducting about 16 sessions per day, assuming everything went perfectly; the sessions would require 24 working days to complete, or about 5 weeks. To reduce the calendar time and meet the deadline, the client accepted our recommendation to assign multiple facilitators.

The Risk We Accepted

Our clients were willing to accept the risk of multiple facilitators to meet their deadlines because, having worked with our consulting firm for many years, they knew us to be extremely structured and careful with consistency in our approach to user studies, and knew we would extend that same care to the multi-facilitator situation. We determined the studies had characteristics that lowered our risk (discussed next), although if we could not meet a threshold of consistency with multiple facilitators, we would be responsible for repeating sessions in off-hours with 1 facilitator to meet the goal.

Our firm had prior experience (in the 1990s) with a study in which we needed to conduct 12 sessions concurrently with visiting sales staff at a client company [1]. For that study, we conducted a 4-hour training session and prepared moderator materials that enabled multiple non-user-experience professionals (client staff) to perform basic facilitation of the sessions. This method is commonly used for game testing as well [2]. In contrast, these studies required that we collect qualitative data at some depth to explain the quantitative data. For that part, we especially wanted senior-level user experience researchers on the team.

CHARACTERISTICS OF STUDIES THAT LOWER THE RISK OF USING MULTIPLE FACILITATORS

The case studies on which this presentation is based posed a lower risk of inconsistency with multiple facilitators, compared to many other types of usability studies that we have conducted. Lower-risk studies for multiple facilitators have the following characteristics. (These characteristics also lower the risk of inconsistency in single-facilitator studies.)

Primary Data is Quantitative or “Multiple-Choice”

When the study questions are answered best by quantitative data (including success rates; number, type, and category of errors; and responses to Likert-scale questionnaires) or pre-determined (tested) multiple-choice answers, facilitators can usually agree on a consistent approach for how to collect the data. In contrast, a study that collects perception data (expectations, reactions, and opinions) is much more subject to facilitator bias and inconsistency across multiple facilitators. Even with quantitative data, opportunity for different interpretations exists that can affect how the data is logged, requiring a consistency-checking step with facilitators as explained further below.

Linear Task Flow

When the task flow participants follow is naturally linear, a natural consistency in the task sequence occurs, creating a common context for user behavior at each step. When the task flow can vary (as is common in most studies), users reach the same points of the process through a variety of paths, introducing variability that increases the complexity of ensuring consistent interpretation by multiple facilitators.

Tasks Familiar to Users

When users are already familiar with the tasks they will perform, the facilitator can give a simple scenario and then observe the user behavior. When users are unfamiliar with the tasks, the facilitator must introduce the scenario with more explanation, some of which may resonate and some which may require further explanation, increasing the complexity of ensuring facilitator consistency.

WHY NOT JUST INSTRUMENT AN UNMODERATED STUDY?

An instrumented study would have enabled us to remove facilitators from the equation altogether. However, both of the case studies sought qualitative data to explain and supplement the findings. An unmoderated study would have required researchers to watch the session recordings to collect the qualitative data, further extending the schedule. What's more, researchers would have had no control over who participated in the study—a critical concern in the Internet marketing study—and no ability to probe in the moment.

COMPLEXITY OF ESTABLISHING CONSISTENCY AMONG MULTIPLE FACILITATORS

Even when studies feature the characteristics that lower the risk of inconsistency when using multiple facilitators, risks abound in the approach.

Facilitation Style and Skill

Different human beings use different vocal inflections, affirmations, and pacing in facilitating research sessions, which may lead to different user responses. In addition, varying levels of skill and expertise with the lab equipment can affect the atmosphere (if in-person sessions) and participant comfort level. Use of facilitators who you know to be extremely skilled and experienced is a necessary prerequisite to reduce variability in facilitation delivery.

Data Interpretation

Different levels of familiarity with the technology or application being studied and different levels of understanding of the purpose of the application and the study goals can lead to different interpretations of the outcomes. If collecting qualitative data is part of the study, different styles of asking users' questions and probing on the answers can increase the risk of inconsistency leading to unnecessary variability. Experienced researchers at least know to take notes about what happened and what was said, versus noting only their conclusions about their observations, so that the notes are consistently the raw data from which to make the interpretation.

TECHNIQUES FOR FACILITATION CONSISTENCY

While it is impossible to make any two human beings act exactly the same, it is possible to bring them into close alignment on how they facilitate a session. The key is having a detailed script, then using practice, script refinement, and trial sessions to bring consistency to what facilitators say, how they say it, and other behavioral factors that otherwise introduce variation.

Script

Many user experience groups confuse a test plan with a script. A test plan is an important first step toward creating a script. It identifies the task scenarios and the questions to ask. What makes a test plan different from a script is that its level of detail allows for and even requires that the facilitator improvise. A test plan does not anticipate possible outcomes and how to handle them. It lacks probing language (and sometimes even task-presentation language).

A script provides all language that the facilitator will utter, as well as "stage directions" for setting up the equipment, resetting the equipment between sessions, and giving the participant materials such as handouts. The language includes reminders of the sequence of probing questions to establish consistency in facilitator intervention; for example, "What are you looking for?", "Take another look.", and "What would you normally do at this point?". Even with a script, facilitators may feel there is some latitude in how strictly they follow it. In a multi-facilitator study, such latitude is forbidden.

Below are snapshots from a script used for the Internet security case study:

Setup instructions (note that sessions were remotely facilitated using WebEx):	**2.1 Before each day's tests:** ▪ Set your screen resolution to 1024x768. ▪ Prepare the backup audio recorder. ▪ Review the first participant's characteristics as captured by the screening survey. ▪ Run the Gordon Shopping-first test module ("Extended Validation Test Module 12_21_06 Gordon first.exe") or Whitley Shopping-first test module ("Extended Validation Test Module 12_21_06 Whitley First.exe"), as appropriate for the day's first participant, and minimize the window. ▪ Start the WebEx session **about 20 minutes ahead of the session start time**. You might still be entering data from the previous participant – that's okay, there's a wait while the participant gets your invite email and connects. 1. Go to http://teced.webex.com/. 2. Click the Log In button at the far right end of the nav bar. The login screen will appear. Enter the username and password given to you and click the Log In button.
Task introduction instructions, which facilitators had to read exactly as written:	**5 Part 1: Introduction to green EV bar and new lock placement [5 min; elapsed 6-8 min]** **Start time**_____ *Maximize the test module with images of IE7 with GordonShopping.com and WhitleyShopping.com:* • *GordonShopping.com has the Extended Validation bar with its company name displayed, which toggles to a line that says "Identified by [CA name]." This site contains a "Click here to learn more" link, which participants can click to read a page about what the green address bar means; in addition, clicking the "Gordon Shopping/Identified by…" section brings up a certificate.* • *WhitleyShopping.com has no green bar or associated live links.* *Press the right arrow to show the Gordon Shopping site.* **>> This is a picture of a fictitious Web site, GordonShopping.com, in the latest version of the Internet Explorer browser, version 7. Version 7 has some new security and privacy features. Click the mouse to take control, then look at the top of the Internet Explorer browser and briefly tell me what security and privacy features you see.**
Key questions we made sure we got right, and how to address not getting a clear answer:	**>> Let's say that you are a regular shopper at GordonShopping.com, and that one day you return and the address bar is no longer green. What would you do?** [ ] Continue shopping ("keep going," "nothing different") [ ] Stop shopping ("abandon," "go somewhere else to find product") [ ] Other: _____ If the participant's response is ambiguous, before using "Other," ask: **>> Would you continue or stop shopping at the site?** If the participant gives an answer that isn't Continue or Stop ("Abandon" is an unlikely answer), use "Other" above. Put answer on Quant tab and notable quote on Quotes tab.

Rehearsal and Script Refinement

Trained user experience professionals typically perform a dry-run session to establish comfort with their script, and then a pilot-test to confirm whether the session as designed works to collect the data of interest. The first few sessions may still feel like practice; by the fourth session, we feel warmed up to our script and can "say it in our sleep."

For multiple-facilitator studies, facilitators must reach that level of comfort before they conduct *any* sessions that will count toward data collection. The facilitators must do more formal rehearsing on their own, in preparation for the trial sessions where final discrepancies among facilitators are identified. A facilitator who comes to the trial sessions without prior rehearsal is wasting everyone's time.

The rehearsal also provides a good opportunity to note suggestions for improving the script. When different people use the same script, they discover phrasings that may be comfortable for one facilitator, do not "roll off the tongue" for another. Rehearsal on one's own, followed by meetings with the other facilitators to discuss and resolve script issues, produces better scripts that the individuals will be more comfortable adhering to, word-for-word. A team lead who resolves differences and updates the script can expedite the process.

Trial Sessions

After learning and rehearsing the script thoroughly, facilitators conduct trial sessions, observed by their fellow facilitators, to receive feedback on facilitation style and staying true to the script. Each facilitator conducts at least two trial sessions, to collect the feedback and then apply it. These sessions can be with stand-in participants such as fellow employees, or even each other. In addition to providing further practice of the script, trial sessions provide the opportunity to practice using the lab equipment, such as making recordings and managing the Web conferencing tool for remote participants and observers.

Eventually, facilitators need to practice with real participants to experience performance anxiety as well as the surprises that real participants introduce, and must be prepared to discard the data if they depart from the script or mishandle an unanticipated event that must be covered in future versions of the script. It's common to go through several real participants before conducting a session that can be counted toward the data collection. Plan to schedule at least two pilot-test sessions per facilitator, and be prepared to discard data from one or two more sessions as necessary.

TECHNIQUES FOR DATA INTERPRETATION CONSISTENCY

The quality of the data collected depends on many factors, one of which is the lens of the researcher who is interpreting it. This interpretation rests on how well the researcher understands the study goals and the issues posed by the application. To ensure that multiple researchers gain the same level of knowledge about the goals and nuances of the application before they facilitate sessions, plan on several things happening:

- Ideally, all facilitators attend all planning meetings and discussions with stakeholders.
- The research team lead documents the details and decisions that emerged from these discussions in meeting notes and distributes them to everyone for review and comment.
- Sometimes not all facilitators are on board when the study begins. In that case, be sure to record all discussions, and require that facilitators added to the team listen to all recordings, read all minutes, and meet with the research team lead to confirm the shared understanding.
- In many studies, data logging may be free-form and done according to the individual facilitator's preference. In multiple-facilitator studies, the structure of the data document is agreed on prior to the sessions. As part of preparing the session materials, the research team lead or another designated researcher prepares a data logging form identifying the elements to capture, and fills in a sample record. The facilitators discuss the elements, refine them, and agree on the format and level of detail of the data to be logged.
- During rehearsal and trial sessions, the facilitators also practice using the data document, review each other's individually filled-in data documents, and identify discrepancies to resolve. Once actual data collection begins, the facilitators log data from early sessions immediately to identify any additional ambiguities with the logging form. For example, we typically found after logging a few sessions the need to split a field into multiple fields for more detailed tagging and analysis.
- In many studies, facilitators use rating scales to record their interpretation of task difficulty, and error codes to classify types of errors. The trial sessions can also establish consistency in use of scales and codes, by having multiple facilitators observe the same trial session and assign ratings and codes to the same activities, then compare their data documents and resolve differences in interpretation.
- Facilitators may continue to consult with one another about user behavior observed during actual sessions for advice on how to log the behavior. Other facilitators might have seen the same behavior previously (or they might in the future).

Here is the error data logging sheet with the available error codes that facilitators needed to use consistently for the search marketing comparison study: It also provided spaces for qualitative data.

7EN	ERROR IN NAVIGATION/INFORMATION	Occurrences:	Step/Page/Path (use ";" to
7EN1	Takes wrong path	2	Change date range/Ad Group page/Clicks Advanced Search and tries to date range;Ad Group page/Clicks Ads tab and ad name
7EN2	Correct path, but then retracts (includes not realizing they were on correct path, return to home or other pages, thinking they're making a mistake when not)	0	
7EN3	Back button repeatedly to get to beginning of steps (just want to start over again)	0	
7EN4	Doesn't scroll to see needed information	0	
7EN5	Unable to find existing information	0	
7EN8	Takes unnecessary/irrelevant steps in completing task	1	
7EN9	Interface does not provide adequate feedback	0	
7EN10	Negates prior completed steps (Re-does a step)	0	
7EN11	Unable to find information using Help link in header	0	
7EN12	Unable to find information using help links/balloons in UI	0	
7EC	**COMPONENT INTERACTION/DATA ENTRY ERRORS**		
7EC1	Attempts to interact with UI element that can't be interacted with	0	
7EC2	Can't figure out how to interact with UI element	0	
7EC3	Misinterprets purpose of UI element	0	
7EC4	Leaves required fields/selections blank (tally for each field) --> Including making a repeat error with the same field	0	
7EC5	Enters incorrect (or improperly formatted) data into field/selection (not including typos)	1	Select keywords/Ad Group page/In Advanced Search, specifies 0 to 2 clicks for the search, which includes 2-click keywords, and doesn't catch he's included them
7EC6	Loses data or has to re-enter information	0	
7ER	**ERROR RECOVERY**		
7ER1	Attempts to proceed without correcting error	0	
7ER2	Performs incorrect action to correct error	0	
7B	**BUGS**		
7B1	(Due to bugs) Has to re-start task from beginning	0	
7B2	(Due to bugs) UI does not function correctly	0	
7B3	(Due to bugs) Has to re-enter field information	0	
7EF	**FATAL ERRORS**		
7EF1	Stops trying or quits	0	
7EF2	Says they would have to call Customer Solutions	0	
7EF3	Believes they've completed the task, but haven't	1	Change bids/Ad Group page/Does not notice he's changed the bid for 2-click keywords

Below is a snapshot of how quantitative error data was coded and tabulated for statistical analysis.

Ppt	1EN1	1EN2	1EN3	1EN4	1EN5	1EN8	1EN9	1EN10	1EN11	1EN12	1EC1	1EC2	1EC3	1EC4	1EC5	1EC6	1ER1	1ER2
A-LK-01	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
A-LK-02	1	0	0	0	1	4	0	0	0	0	0	0	0	0	0	0	0	0
A-LK-03	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0
A-LK-05	6	0	0	0	0	0	0	0	0	0	1	0	0	0	2	0	0	0
A-LK-06	4	0	0	0	0	2	0	0	0	0	0	0	0	1	1	0	0	0
A-LK-07	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0
A-LK-08	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
A-ME-01	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
A-ME-02	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A-ME-03	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
A-ME-04	1	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
A-ME-05	9	0	0	0	0	3	0	5	0	0	0	0	0	0	0	0	0	0
A-ME-06	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
A-ME-07	5	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0
A-ME-08	4	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0
A-ME-09	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A-ME-10	3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
A-ME-11	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A-SS-01	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
A-SS-02	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	0	0	0

RESEARCHER QUALITIES THAT LOWER THE RISK OF A MULTIPLE-FACILITATOR SOLUTION

A researcher who is a good fit for a multiple-facilitator study must combine the qualities of a rigorous researcher and a team player:

- Proven experience adhering to a script rather than the "winging it" style of facilitation
- Rigorous attention to detail:
  - Willingness to record details of what happened in each session (participant clicked this button, entered that keyword, made a face, uttered this comment, etc.), not just a higher-level interpretation of the observations (typical path, no real problems)
  - Persistent focus and patience when interpreting the detailed data, including ability to recognize and correct discrepancies in one's previous logging
- Willing to do the additional practice needed for a multiple-facilitator study
- Comfortable with critique and negotiation
- Ruthless devotion to consistency: willing to discard session data (for example, after an unsuitable participant or technical difficulties) or to admit a script deviation and discuss solutions with team members

CONCLUSION

In large studies with challenging schedules, the multiple-facilitator approach is a viable alternative to long session days for a single facilitator, which can lead to fatigue and intra-facilitator inconsistency. The keys to establishing consistency in multiple-facilitator studies are teamwork, practice, and structured data collection.

REFERENCES

1. "Multiple-User Testing: When One Person Can't See Everything" (by Laurie Kantner, Proceedings of the 1997 Usability Professionals' Association Conference, also found at http://www.teced.com/PDFs/upa97lk.pdf).
2. "Multiple-User Simultaneous Testing (MUST)" (by Jakob Nielsen, Alertbox, October 15, 2007, found at http://www.useit.com/alertbox/multiple-user-testing.html).

ABOUT THE AUTHORS

Lori Anschuetz, Senior Consultant
TecEd, Inc.
lori@teced.com

Lori Anschuetz has over 30 years of experience as a user researcher, information architect, and project manager, including more than 20 years at TecEd. Lori's usability evaluation experience includes home visits and laboratory tests, as well as expert evaluation, user interviews (in person and remotely facilitated), focus groups, contextual inquiry, and ethnographic interviews. She has also been one of multiple facilitators on large-scale usability studies involving 60 to 400 participants, and has worked with users throughout the United States, as well as in Canada, the United Kingdom, France, and Germany. She is a member of UPA and Western New York HFES, and a senior member of Society for Technical Communication (STC).

Laurie Kantner, Former VP of Client Services
TecEd, Inc.

Laurie Kantner has been a UX practitioner since the early 1990s and has defined and carried out scores of user research projects over her career, from small to extremely large, using lab and field methods, evaluating software, Web applications, services, and devices. Her earliest work in the area of multiple facilitators was presented at UPA 1997 in the talk titled "Multiple-User Testing: When One Person Can't See Everything." Laurie is a member of the Usability Professionals' Association, ACM SIGCHI, and the Society for Technical Communication (STC). She has contributed chapters to *Essays on Usability: Design by People for People* (UPA, edited by Russell J. Branaghan) and *User-Centered Design Stories: Real-World UCD Case Studies* (Morgan Kaufmann, edited by Carol Righi and Janice James).