

Jane Nelson ©1997 Artville, LLC

Penny-Wise, Pound-Wise: Making Smart Trade-offs In Planning Usability Studies

When you are doing usability or other user studies, reality always imposes constraints—on time, on personnel, or on money. The issue is not whether to make trade-offs, but, rather, how to make them so that your usability efforts are truly cost-effective. Beware of expediency as a basis for decision making. We have seen situations where cutting the wrong corners in the interest of sometimes relatively small, apparent savings undermined the value of the rest of the sizable investment in a project. You don't want the value of your study compromised by skimping on one area, especially if other trade-offs might have yielded bigger savings.

Where, then, can you depart from the “ideal” and still have a project that is feasible within your constraints and yields the valuable information you are seeking? As always in human factors, “it depends,” but here is a rough guide based on what we have seen in working with many clients doing user studies. This article does not pretend to be the result of a scientific survey, or of a comprehensive literature review, but rather is based on our experience in the field, including some lessons learned in the school of hard knocks.

Dangerous Trade-offs

RECRUITING

Efficient use of your personnel time, and good return on your “fixed costs” such as facility rental, require a successful recruiting effort. The quality of the resulting data depends on the degree to which the people you are studying reflect the target group of users. Spend the money it takes to get the right people. If you are pleasantly surprised by a lower-than-expected bid for recruiting—be cautious. Also, do not skimp on resources planning and overseeing the recruiting process.



Selection of Evaluators

In our experience, good recruiting requires a well-developed and thought-out screener. This does not necessarily mean a highly elaborate screener. We have seen problems with screeners that are too broad, too narrow, or too complex.

Too Broad

Sometimes, in an effort to ensure an “easy” recruit, the screener may be made too broad. This is more likely to happen if you are pushed into a quick recruit because of not having allowed enough time or other resources to support the recruit. The risk, of

course, is that you won’t screen out people who are unlikely to be users of your product.

For instance, if you have a broad screener for a usability evaluation of a software product, you may have people who have never used a particular platform, or a particular category of product. If your strategy is to design an entry-level product that requires no prior knowledge of hardware or software a broad screener may suffice, but if this is not your strategy, the result can be disastrous. The confounds are significant when, for example, someone has to install software in a usability evaluation, but has never done so, and would never do so in real life. Are any problems you find attributable to a user’s inexperience or to problems with your product? It is impossible to tell unless that is the only problem he has—which is unlikely.

Another problem with a broad screener is that you may not be able to make any generalizations if you have a range of responses or see a variety of situations (as in a visit study). Is the difference due to real differences or is it an artifact of the breadth of the sample?

Too Narrow

Conversely, if your sample of evaluators is too narrow, you limit the ability of your results to be general. Again, if this matches your product strategy, this is not a problem. For example, if you are working on a highly specialized product for a particular audience, you need a highly targeted sample. Be careful that your sample matches your product strategy. This is a common problem when you have to rely only on internal people in your own department (often with no formal recruiting at all) to test a software user interface, for instance. The fact that software designers find your interface intuitive is probably of little help in predicting whether it will be acceptable and usable by the public at large.

Too Complex

A screener that is very complex or very long can also compromise a study. Complex scoring algorithms are rarely required, may make it difficult to determine what population you are really sampling, and may simply intro-

duce practical problems. In one study we did, the screener supplied by the client was so complex that recruiters had to write down the information, have the manager calculate the score, and call the prospect back to arrange for the actual testing. In many cases, people who would have otherwise qualified were not included because the recruiter could not reach them a second time, or because they refused to come, in some cases citing the interview on the phone as the reason they didn't want to participate.

Recruiting strategies often need some adjustment after their initial results, which requires a commitment to actively manage the process. If a too narrow screener is yielding too few participants, it may need to be broadened, but judiciously. Do not skimp on oversight and monitoring of the recruiting process. It is definitely not something to hand over to a market research company and turn your back on.

Budgeting Time for Recruiting

Good recruiting takes time. Take the advice of the recruiter about how long it takes to recruit someone. Giving him less time will result either in his not being able to get enough people or, worse, getting the wrong people. Too few people is less of a problem if you are doing a local, or internal test, because your other investment will be lower. But if your test involves expenses for national or international travel, if you have committed to a reservation at a facility, or you have other fixed costs, you certainly want the study time to be maximized. Allow sufficient time to adjust your recruiting strategy if necessary.

Paradoxically, recruiting too far ahead also can be a problem—the drop-out rate increases as other things arise. If you are going to recruit far ahead, be sure to maintain contact—or reconfirm about a week before testing (and again the day or night before if possible).

Incentives

We all want to get the right people for as little money per person as possible, but if your incentive is too low, you may not get the right

people. For some recruits, you must expect to pay significant incentives. For instance, to recruit network administrators, highly placed technical people and their managers, or “experts” in almost any field, typically requires high incentives. If you try to cut corners here, you may be able to locate the right people but not get them to agree to participate. Worse yet, they may agree to come, but fail to appear, resulting in most of the expense with no data.

FACILITATOR SKILLS

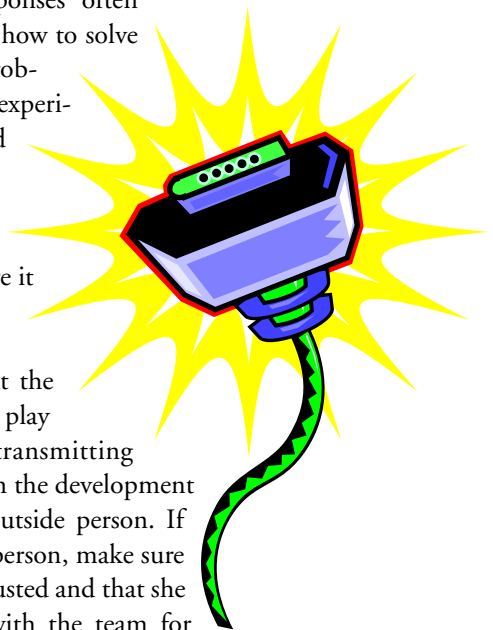
A good facilitator is key to a good study. If your facilitator leads the evaluator inappropriately (or without being aware of doing so), is not versed in the art of open-ended questions, fails to follow up on behaviors of interest, does not establish good rapport with the evaluator, or does not address the questions from the project team, the evaluation can be compromised. The facilitator must understand the design issues well enough to know where to probe and be experienced enough to judiciously depart from or elaborate on the prepared protocol.

Remember also that the facilitator's own impressions are a significant source of data. A good deal of information comes from the facilitator's first-hand impression of subtle cues in the interaction. Not only does this help identify problems, but the close observation of the user's responses often contributes clues about how to solve identified usability problems. All of this takes experience and specialized skill.

Team politics may also make facilitation an area where it is important to bring in an outside person. Because of the role that the facilitator can or should play in interpreting and transmitting findings, credibility with the development team may require an outside person. If you do use an internal person, make sure that her objectivity is trusted and that she has enough standing with the team for



Susan M. Dray
David A. Siegel
Dray & Associates, Inc.
2007 Kenwood
Parkway
Minneapolis, MN 55405
+1-612-377-1980
fax: +1-612-377-0363



her input to be incorporated.

Obtaining the right mix of facilitator skills can be a real challenge in international testing. It is better to use a trained usability person who speaks the language, even if not perfectly, rather than a local person who is fluent, if they are not skilled in interviewing or open-ended questioning, or if they are not willing to work with a trained usability person to develop an appropriate style of probing. In some situations, the needed mix of usability and language skills means you will need a local speaker paired with an experienced usability specialist.

PLANNING THE FOCUS OF THE STUDY

Although it may seem an added up-front cost, beginning a usability evaluation project with a thorough review of the system or product and of user and usage information can greatly facilitate protocol development and result in a more cost-effective study overall. Although this type of review may resemble a traditional expert review, it is not aimed at a detailed critique of the design and at specific recommendations for changes. Rather, it is focused on prioritizing the design issues and user tasks where formal evaluation will likely be most useful. With even a moderately complex product or system, it is rarely feasible to evaluate every functionality or navigational pathway.

PROTOCOL

Failing to set the stage adequately or accurately, providing too much or too little information, or not having a good idea of the key areas to be probed are all classic mistakes. It is hoped that a good facilitator can help identify where these potential problems will occur, but even a good facilitator cannot make up for a poorly designed test protocol. This is definitely not a place to cut corners. Spend the time as a team to identify key tasks, scenarios that will tap the tasks, the information required for

successful completion of these tasks, and a logical order or flow. If some tasks logically depend on successful completion of others, be sure to prepare “dummy information” so you can still have evaluators attempt later tasks even if they are unable to complete earlier ones.

TEAM INVOLVEMENT

Make sure that there is adequate team involvement in planning the study and in participating in the tests. This is important to ensure that the study really does meet the team needs and that its results will be taken seriously in the design process. Obviously, it is “easier” and “cheaper” to have a usability specialist, whether internal or external, go off and do the study on his or her own. But the data is much less likely to be useful or used.

Lower Risk Trade-offs

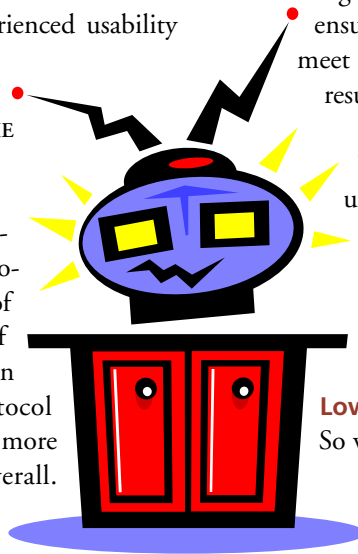
So where can you make trade-offs?

Facilities

Sometimes you do need to rent a full usability lab or a focus group facility. If you have a large audience to which you can introduce the concept of user testing, a highly political project where buy-in to the results is critical, or a sophisticated setup that requires pan and tilt, zoom, scan conversion, and the like, you may need to pay for lab rental. However, often this is an area where the cost is high and the payoff is not really justified. Sometimes a conference room with the design team in the corner, or a video camera on a tripod in the corner with the team watching a monitor in the next room, is equally effective and much less costly. Often, on-site or naturalistic testing may be appropriate, eliminating the need for facility rental.

Videotaping

Most videotapes are never looked at again. It is time consuming to review them, and creating an edited version can take twice as long as the initial testing, or more. Sometimes, of



course, videotapes are critical. For instance, tapes are useful early in establishing a usability program in a company, or at other times, such as when a key team member or manager is unable to attend the evaluation, when a highly charged political issue is being investigated, or when, as in the case of international evaluations, most of the team is unable to attend.

Of course, if you know in advance that only an edited composite tape will do, you know that you must video tape. However, the added expense of an edited composite is not always called for. The same purpose can sometimes be achieved by cueing up your tapes to the intended spot. This can significantly reduce the time, since no rerecording is required.

Scripts

Although I have emphasized the importance of protocol development, an exactly worded script is rarely needed or useful except for the most novice facilitator. Even for that person, the variability of human behavior makes a formal script unlikely to be useful. Scripts can be a hindrance. Better to spend the time and effort on training the novice in how to “follow” an evaluator, how to let him make and recover from mistakes, how to question without leading, and what to do when evaluators become frustrated or unable to continue than to create a formal script to be memorized and followed.

Number of Evaluators

Some usability evaluations do indeed call for large numbers of evaluators. However, numerous studies have suggested that before you reach 10, you have reached the point of diminishing returns.

Of course, if you have clearly differentiated groups, or different levels of functionality that will be accessed by distinct types of users, you may need to have multiple sets of evaluators. However, you should carefully think about whether you really need to distinguish a “critical group.” We have seen the costs for studies skyrocket as people began to define somewhat speculative groups and push for multifactorial

research designs. Remember that introducing new comparisons increases costs geometrically.

A corollary to this is that a decision to test in multiple geographical locations should not be made reflexively. One practical reason for going to multiple locations is that the product may be specialized, and the population of users in any one area is too thin for adequate sampling. However, using multiple geographical locations is usually a priority only if you have reasons to suspect real demographic, contextual, or usage pattern differences likely to effect usability. With products intended for a broad international market, this is more likely to be a factor, but less likely with products intended for a national market or products that are building on a well-established platform. Designs that are highly innovative may call for evaluation of fundamental design choices that can be assessed with any fairly representative local group of users—good news for start-ups that often produce such products and have limited resources for usability evaluation.

Statistics

Similarly, just as testing rarely requires complex research designs, most of our evaluations have not called for more than simple descriptive statistics. Often even these are not necessary because of the power of the qualitative observations. Even in cases where interest might exist, the numbers of evaluators are rarely sufficient for meaningful statistics, and the time required to compute them is not well spent. It is more common that we find major usability issues that would have been difficult for the team to anticipate, but that “leap out at you” when you see users struggling with them and don’t need to be teased out through statistical analysis.

Formal Reports

As with videotapes, formal reports often go unread. In some organizational settings or

BUSINESS COLUMN EDITOR

Susan Dray

Dray & Associates, Inc.

2007 Kenwood Parkway

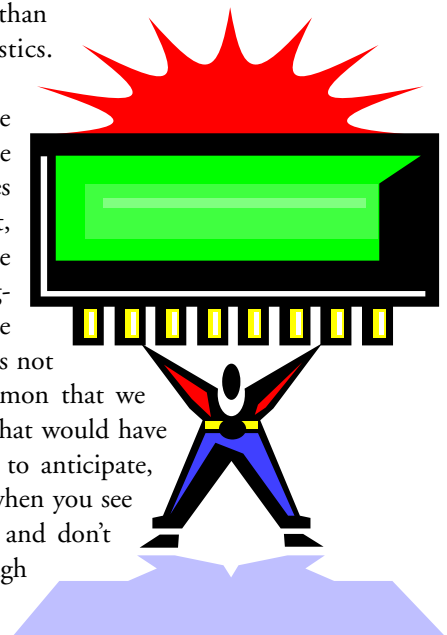
Minneapolis, MN 55405,

USA

+1-612-377-1980

fax: +1-612-377-0363

dray@acm.org





with some types of products involving liability issues, a formal report may be necessary to document findings. However, usually the important thing is simply that findings be meaningfully conveyed to the team to influence the design process. It is often less expensive and more effective to use informal, participative approaches to transmit findings and involve the team in examining them. These can be supported as needed by data summaries, vignettes, and other techniques for displaying qualitative information.

Conclusions

Some generalizations emerge from putting together all of these recommendations. Overall, you will derive more benefit from your usability dollars by doing studies that are simple in design but adequately supported for

planning, recruiting, and usability skills such as human factors knowledge and test facilitation. Furthermore, good planning has to do more with developing good communication, understanding of the issues, and mutual confidence between the development team and the usability expert than with developing rigid verbatim protocols and scripts (assuming adequate usability skills). This mindset is certainly easier to achieve when usability evaluation is iterative, beginning early in the process, than when it is put off until the end, waiting for the finished product to be ready to test and the perfect study designed. Of course, taking this one step further brings us to an oft-cited theme—the most effective use of usability dollars is to build usability into the development process from the beginning, rather than treating it as a major add-on at the end. ☺

PERMISSION TO MAKE DIGITAL OR
HARD COPIES OF ALL OR PART OF THIS
WORK FOR PERSONAL OR CLASSROOM
USE IS GRANTED WITHOUT 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 SERVERS OR TO REDIS-
TRIBUTE TO LISTS, REQUIRES PRIOR
SPECIFIC PERMISSION AND/OR A FEE.
© ACM 1072-5220/99/0500 \$5.00