University of Texas at El Paso
DigitalCommons@UTEP

Departmental Papers (CS) | Department of Computer Science

10-22-2007

Toward a more accurate view of when and how people seek help with computer applications

David G. Novick
University of Texas at El Paso, novick@utep.edu

Edith Elizalde
The University of Texas at El Paso

Nathaniel Bean
The University of Texas at El Paso

Follow this and additional works at: http://digitalcommons.utep.edu/cs_papers

Part of the Computer Engineering Commons

Recommended Citation
Novick, David G.; Elizalde, Edith; and Bean, Nathaniel, "Toward a more accurate view of when and how people seek help with computer applications" (2007). *Departmental Papers (CS)*. Paper 17.
http://digitalcommons.utep.edu/cs_papers/17

This Article is brought to you for free and open access by the Department of Computer Science at DigitalCommons@UTEP. It has been accepted for inclusion in Departmental Papers (CS) by an authorized administrator of DigitalCommons@UTEP. For more information, please contact lweber@utep.edu.

Toward a More Accurate View of When and How People Seek Help with Computer Applications

David G. Novick
Department of Computer Science
The University of Texas at El Paso
El Paso, TX 79968-0518
+1 915-747-5725
novick@utep.edu

Edith Elizalde
Department of Computer Science
The University of Texas at El Paso
El Paso, TX 79968-0518
+1 915-747-5480
eelizalde@utep.edu

Nathaniel Bean
Department of Psychology
The University of Texas at El Paso
El Paso, TX 79968-0523
+1 915-747-5000
nbean@utep.edu

ABSTRACT
Based on 40 interviews and 11 on-site workplace observations of people using computer applications at work, we confirm that use of printed and on-line help is very low and find that providing greater detail of categories solution methods can present a more realistic picture of users' behaviors. Observed study participants encountered a usability problem on average about once every 75 minutes and typically spent about a minute looking for a solution. Participants consumed much more time when they were unaware of a direct way of doing something and instead used less effective methods. Comparison of results from different data-collection methods suggests that interviews, and probably surveys, provide less reliable views of users' problem-solving behaviors than do participatory evaluation and direct observation.

Categories and Subject Descriptors
H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *Evaluation/methodology, training, help, and documentation.*

General Terms
Human Factors, Measurement

Keywords
Documentation, usability, evaluation

1. INTRODUCTION
Research in the last ten years has shed new light on when and how people seek help in handling problems they experience in using computer applications. However, there are still big gaps in the picture of people's use of documentation and help systems in practice. These gaps are largely the consequence of known methodological issues, but the methodological difficulties of understanding actual user behavior in the workplace are significant enough that prior studies have used methods that, even if flawed, nevertheless generated otherwise unobtainable results.

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, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
SIGDOC'07, October 22–24, 2007, El Paso, Texas, USA.
Copyright 2007 ACM 978-1-59593-588-5/07/0010...$5.00.

One of the focuses of recent research in this field has been users' choice of methods for solving problems they encounter with computer systems. In this paper, we address two issues in particular: lack of detail within categories of solution methods (such as seeking on-line help) and the validity of the methods used in this research to determine how often users employ these various kinds of solution methods.

1.1 Usability and Documentation
Despite the efforts of researchers and practitioners in an entire subfield of computer science, usability problems still plague information technology. Much has been published on how to find and eliminate usability problems in computer applications (e.g., [3], [6], [5], [15], [20]), and various approaches to detecting usability problems have been compared for effectiveness (e.g., [5], [9], [14]). Yet usability problems still crop up, and users still experience lots of problems in practice [1], [4], [7], [12], [16]. And because these problems typically arise from some aspect of the context of use, eradicating all usability problems may be unrealistic.

A practical consequence of unresolved usability problems is that users must be provided training and support in the use of computer applications. The cost-effectiveness of these efforts depends on whether the users actually refer to and benefit from these forms of support, particularly beyond the start-up phase of novice errors (c.f., [10], [12]). Thus, both for users and publishers of computer applications, much depends on the actual extent of usability problems and the actual use (and usefulness) of documentation and help systems. By extension, these stakes also depend on the effectiveness of methods for determining actual usability problems and actual use and usefulness of documentation and help systems. Otherwise, users are making choices about software and developers are making choices about support of software that have impacts of many billions of dollars while relying on little more than educated guesses.

1.2 Impact of Methodology
The research into the incidence of frustrating problems with computer systems and how people address these problems has difficult methodological issues, stemming largely from the nature of the phenomena being studied. As we discuss in Section 2, the answers are hidden in plain sight: the behaviors are everywhere around us in the workplace but hard to collect and assess. As a consequence, the views we have of use of documentation and help systems are inconsistent. Some studies (e.g., [17], [11]) have found relatively high levels of use of documentation and help. Other studies (e.g., [4], [12]) have found stunningly low levels of

use of documentation and help. Other studies (e.g., [16]) present a more mixed picture. Which of these results is correct? In this paper, we argue that the usefulness of a view into actual use depends on the methodology used in the particular study. And some methodologies for assessing use of documentation and help systems appear to be more effective than others. How accurate are contemporaneous self reports? How accurate are data obtained from interviews?

To address these issues, we review the methodological issues arising from recent research into how users respond to usability problems, we explain this study's approach to filling in gaps in the prior studies and assessing the methodological effectiveness of their findings, we present the study's findings from 40 interviews and 11 on-site workplace observations, and we conclude with suggestions for next steps in understanding when and how people seek help with computer applications.

2. RELATED WORK

Research into how users of computer applications solve usability problems has included self-reports, Web-based surveys, telephone interviews, and in-person interviews. The results vary. Some studies indicate that use of documentation is widespread [17], [18], but this was based on a single application, largely among novice users. Other studies indicate that people rarely use on-line help and almost never use printed manuals [4], [12]. The studies' participants used on-line or printed documentation in only 0 to 4 percent of the occasions they encountered problems with their use of computers. These studies, though, were based on self-reports and peer observations from relatively homogeneous populations, such as computer science students [3] or middle-school teachers using a single application [5].

A more recent study [14] assessed use of documentation across a much more heterogeneous sample of computer users. This study found that, on average, participants estimated that they used on-line help in about 28 percent of the occasions in which they experienced difficulty and used printed documentation in about 3 percent of these occasions. Indeed, more participants reported that they abandoned a task than used printed documentation. But, as the authors pointed out, this study, too, had methodological weak points. In particular, the interviews to assess solution patterns may have been unreliable, and. some solution categories were probably too broad. Concerns about the interview methodology relate to the validity of participants' accounts of their use of documentation. The issue with the solution categories arose because the category "asked other" could cover asking a colleague, asking someone at an internal or contract help desk, or asking someone at the software publisher's help desk. The category "used on-line help" could cover the help provided with the application, help available from the publisher via the Web, and help available from unofficial sources such as online forums and newsgroups, usually located via a search engine. The interviews suggested that the study's participants considered these to be different sources of information. Similarly, the category "used printed manual" could include both the manual supplied with the software or an "after-market" book.

Our review of related work suggests that methodological issues posed, and continue to pose, problems for research into the incidence of and users' responses to usability problems. These issues include:

- *Novice users*. Some of the studies [18], [17] looked primarily at novice users. This approach presents a number of problems. First, novice users are in the learning phase and thus may naturally refer to documentation more often than experienced users. Second, novices tend to encounter different kinds of problems than experienced users [12]. And third, the data reported by [16] may indicate that users refer to printed documentation only when installing a new application. For these reasons, studies focusing on novice users may overestimate the extent to which people use documentation.
- *Unrepresentative users or work context*. While any user is representative of himself or herself, the studies showing the lowest rates of use of documentation were based on special groups of users or work situations, rather than sampling broadly in the workplace. This likely occurred because of difficulties in obtaining broadly representative subjects or in finding subjects who could be studied longitudinally with a common application. Thus the participants in [4] were computer science students, the participants in [12] were teachers at a middle school, and the participants in [11] were members of the public, including members of underrepresented groups, blind users, and developers and technical support providers who were willing to answer a Web-based questionnaire of about 50 questions with compensation of a $10 gift certificate. The methodological issue is that neither computer science students nor middle school teachers may have the kinds of usability problems and responses that are characteristic of the broader world of work. And participants willing to fill out a long Web form may not represent the general population of users of computers, either. Indeed, the methodological similarities and disparate results of [4] and [12] suggest that subject population and work context plays a large role in the outcome of the study.
- *Limited software*. Some research examined use of documentation and help systems for a single application. One study [18] focused on a particular word-processing application, and the researchers obtained subjects through a list of purchasers provided by the publisher. Another study [12] looked at use of a single module of a software package for teachers. Until the field has many more such studies, we do not know if the observed phenomena are generally true or rather limited to the circumstances of the particular application.
- *Interviews*: Whether conducted face to face or by telephone, interviews enable researchers to go into depth but have limited validity. The problem is that people remember things in ways that are systematically skewed. For example, when asked to remember the most recent incident of a certain type, they are likely to remember the most salient rather than the most recent [19], [2]. Thus studies of use of documentation and help systems that relied on interviews (e.g., [18], [16]) may have results that are skewed toward salience rather than recency. And, as we discuss in Section 5, interviews about documentation and help systems may also suffer from other systematic problems, such as social desirability bias and fatalistic acceptance of usability problems with software.
- *Participative evaluation*. In contrast to conducting post-hoc interviews, some studies [4], [12], [13] have relied on contemporaneous reports or diaries from study participants. This technique, called participative evaluation [8], has fewer problems with recall, but does not permit the researchers to

obtain clarification or go into depth because they are not present to ask additional questions of the participants when warranted. Additionally, as we discuss in Section 5, participative evaluation misses cases where participants mistakenly believe that they do not have a usability problem.

- *Surveys*. Although they may facilitate including larger numbers of participants and may include dozens of questions (cf., [11]), surveys suffer from the ills of both interviews and of participative evaluation. That is, they are subject to recall effects yet do not allow researchers to follow up answers with clarification.
- *Categories with insufficient detail*. Since the lead study by Ceaparu et al. [4], subsequent studies of frustration with computers have tended to use similar categories for the participants' solutions (or non-solutions), which permitted the researchers to compare results. The interview methodology of Novick and Ward [16] enabled going into greater depth when the participants described their attempts to solve usability problems, and this disclosed that at least three solution categories—"asked other," "used on-line help" and "used printed manual"—might be overbroad. For example, using "on-line help" includes both using the help supplied with the software and visiting a Web-based forum.
- *Observation*. While observation of subjects provides the most direct view of users at work, this advantage comes at considerable cost. Participants and their employers can be reluctant to permit observation in the workplace, particularly if audiovisual recordings are made of the participants at work (The observations reported by Ceaparu et al. [4] were made by student peers). Clarification of behaviors is difficult to obtain because questions from the observer would interfere with and thus change the participants' activities. Observation is time-consuming because in practice frustration episodes turn out to be relative infrequent; a two-hour session may not include any frustration episodes. And the clearer picture of the participants' work brings with it questions not faced in other methods, such as coding of tasks in which the participant uses a work-around but is unaware of the existence of the better technique.

Given these methodological issues, in this study we address these questions:

- Can possibly overbroad solution categories be clarified?
- How reliable are interviews as a methodology for assessing users' ways of solving problems with computer applications?

3. METHODOLOGY

In this study, we were specifically interested in clarification of categories of solution approaches and in the nature of the relationship between interview reports and observed behaviors with respect to usability-problem episodes and solutions tried. We expected that the use of self-reports through interviews might lead to systematic biases in the kinds of solution approaches attempted. Thus the study had two phases. The first phase replicated Ward and Novick, conducting interviews with more subjects and with more detail, in order to clarify possibly overbroad solution categories. The second phase returned to a subset of the same subjects, conducting in-person, on-site observation of the subjects at work.

3.1 Interviews

The interview phase included 40 participants, comprising 20 women and 20 men. The average age of the participants was 41.8. As indicated in Figure 1, most of the participants had at least some college. And as indicated in Table 1, the occupational distribution of the participants was reasonably broad, with particular representation from managers and professionals, who are particularly likely to be conducting work using a computer. We consider 7 of the 40 subjects to have a high level of general technical proficiency.

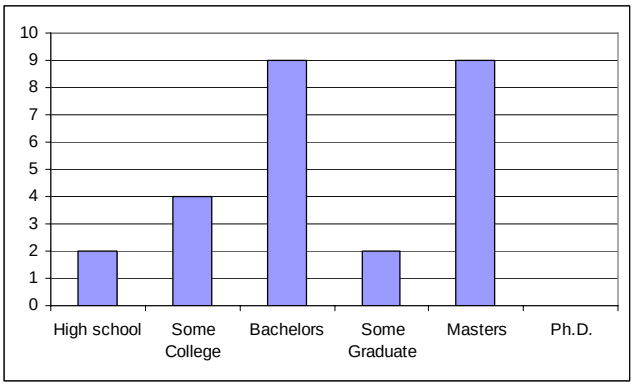

Figure 1. Number of study participants per level of education.

Participant Occupation Category	Number
Management	14
Technical	9
Public Relations	2
Assistant	4
Military	1
Service	3
Professional	7

Table 1. Distribution of Occupations.

The participants were recruited and the interviews conducted using methods approved by Institutional Review Board of the University of Texas at El Paso. Subjects were not compensated for their participation. The interviews covered demographic information, operating system and applications used, frustration episodes with the applications, solutions, and self-estimated distributions of solution approaches. In asking for these distributions, the researchers provided possible categories that included detailed choices within the general categories classified in earlier studies as "asked other" and "used on-line help."

As was the case in [16], the participants overwhelmingly used Windows as their operating system. And the Microsoft Office applications Word, Excel, Outlook and PowerPoint were the applications reported as most frequently used by the participants. Participants reported relatively few problems with database applications (probably because relatively few of the participants used databases), but these problems led to higher levels of frustration than for other applications. PowerPoint had the least-frustrating problems, and participants reported no frustration episodes involving Web browsers. Proficiency levels and frustration levels did not appear to be correlated. Nor did the number of reported episodes and frustration levels appear to be correlated.

3.2 Observations

The observation phase was conducted with the 11 of the 40 interview subjects who agreed to permit one or two of the researchers to spend about two hours observing them at work. The subjects included eight women and three men. The researchers' notes of the participants' behaviors and interactions were recorded on laptop computers with automatic time-stamps for entries. The researchers recorded the participant's applications, tasks (as best could be determined), problems with the use of the computer, what the participant did, if anything, to address these problems, and whether this approach was successful. In noting problems with the computer, the researchers included not only problems recognized by the participants as such but also other problems they did not recognize, such as using a high-effort approach where the application provided a simple method. For example, one participant laboriously produced individualized documents, when she could have used a "mail-merge" function provided by the application.

The work places varied from a blood bank to a chamber of commerce. In all, the researchers produced transcripts of about 22 hours of the work lives of the participants. The transcripts were then coded for usability problems and solutions, and the coded data were summarized, compared to the data for the interviews, and explored for qualitative insights.

4. RESULTS

4.1 Interviews

The results of the interview phase suggest that the overall pattern of solution methods reported by users of computer applications is stable. We asked each of the participants to estimate the percentage of the times they tried various kinds of solutions when they encountered frustrating problems in using computer applications. Figure 2 compares the distribution of solution approaches reported by subjects in this study with the distribution reported by Novick and Ward [16]. The data for the current study include combined data for the broad categories of "asked other" and "used on-line help." The solution distributions of the two studies have a correlation coefficient of 0.90.

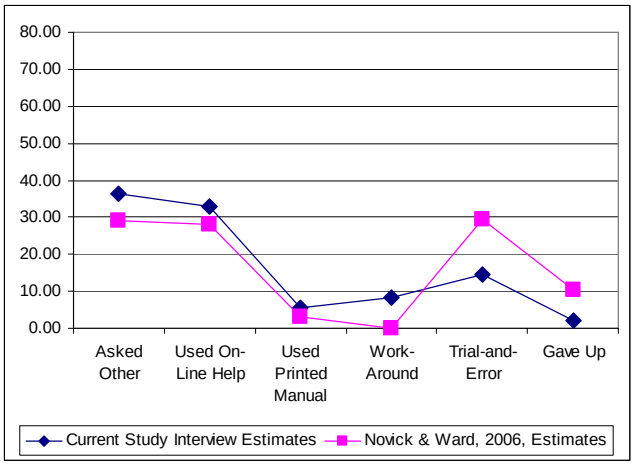

Figure 2. Mean reported solution methods.

Looking at the increased detail for the categories of "asked other" and "used on-line help," it turns out that additional categories can improve our understanding of on-line help and asking others but are less helpful for the category of using a printed manual. As indicated in Figure 3, the broader category "asked other" does break out into two roughly equal sub-categories, "help desk" and "colleagues;" their respective means are 19.55 percent and 16.92 percent.

Similarly, the category "used on-line help" breaks down into multiple sub-categories. As indicated in Figure 4, both local on-line help and search engines figured prominently in users' reported solutions to usability problems. Participants on average reported using local on-line help 20.66 percent of the time and using a search engine 8.98 percent of the time. However, users visited chat sites and manufacturers' Web sites 0.80 percent and 2.24 percent of the time, respectively. While the categories appear to vary considerably in their reported use, the fact that use of a search engine accounts for nearly 9 percent of solutions suggests that good methodology would involve use of detailed solution methods within the general category of "used on-line help." In contrast, as indicated in Figure 5, use of aftermarket printed materials appears to be low enough relative to manufacturers' manuals that this category can be left as is. However, the meticulous researcher may choose to include both sub-categories.

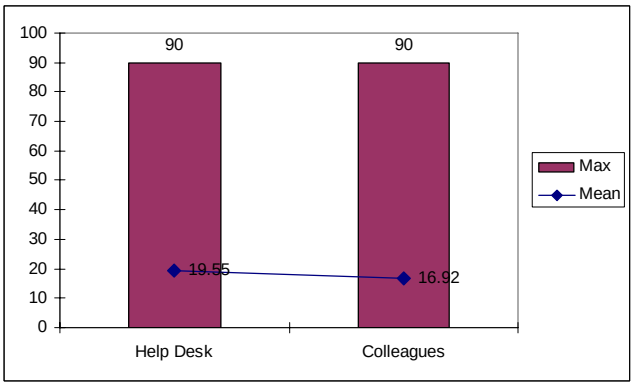

Figure 3. Mean and maximum percentages of reported solutions within the "Asked other" category.

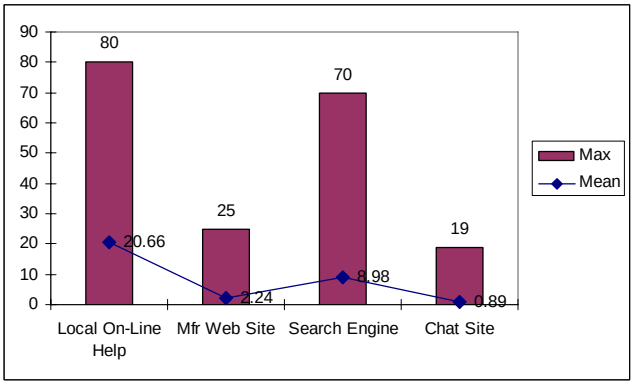

Figure 4. Mean and maximum percentages of reported solutions within the "Used on-line help" category.

Comparing key findings with those of Novick and Ward [16], we note that estimates of last use of printed and on-line documentation are particularly consistent, as can be seen by comparing Figures 6 and 7. To facilitate comparison, Figure 6 does not include data where the participant either did not remember their last of documentation or said that he or she had never used documentation. The data may be smoother in the present study because more subjects, 40 rather than 25, were interviewed. The correlation

coefficient between the results of the current study and those of Novick and Ward for last use of on-line documentation is 0.95.

We note that the data for last use of on-line help are consistent with the results reported by Martin, et al. [11], who found that 20 percent of users of PC applications sought on-line help daily and that another 39 percent sought on-line help at least once a month. And although visually a correspondence seems evident, the correlation coefficient between the results of the current study and those of Novick and Ward for last use of printed documentation is 0.31.

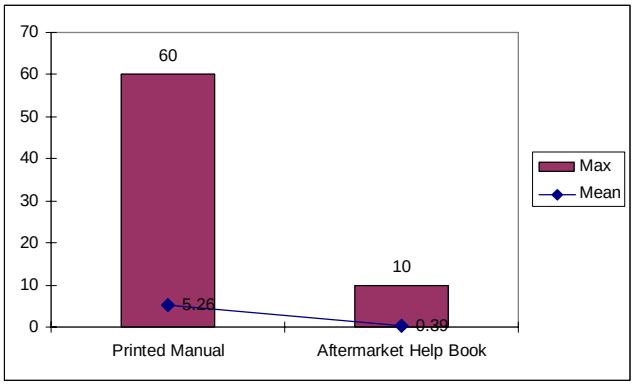

Figure 5. Mean and maximum percentages of reported solutions within the "Used printed manual" category.

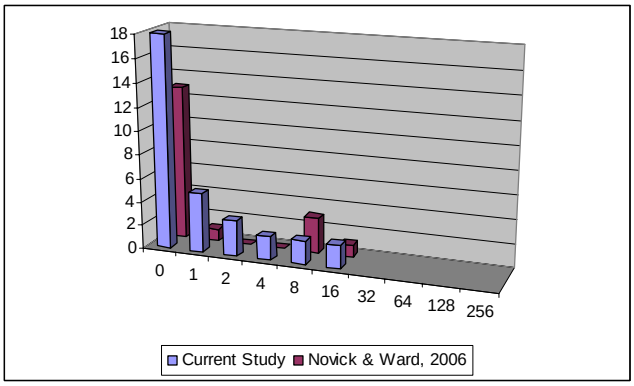

Figure 6. Comparative histograms, months since last use of on-line documentation, exponential scale.

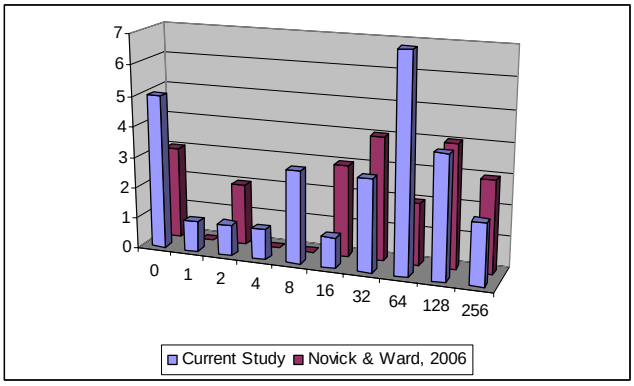

Figure 7. Comparative histograms, months since last use of printed documentation, exponential scale.

We speculate that the bimodal distribution of distribution of last use of printed documentation—there are distinct maxima at 0 and 64 months—reflects that some participants still use printed documentation while many others look at the documentation when they start using an application and then never look at it again. Perhaps the participants used the documentation for installation or read a tutorial, and thereafter just use other forms of help, if any.

4.2 Observations

We observed eleven of the interview participants for about two hours each. In the 22 hours of work we observed, we noted 16 frustration episodes, which were distributed among the subjects as shown in Figure 8. Typical usability problems leading to user frustration included formatting text, saving files, sorting and summing data, inserting text and links.

The mean time of the frustration episodes (from noticing the problem until the user solved it or gave up) was 1.8 minutes, with a standard deviation of 2.0 minutes, and a median of 1.0 minute. Figure 9 shows the distribution.

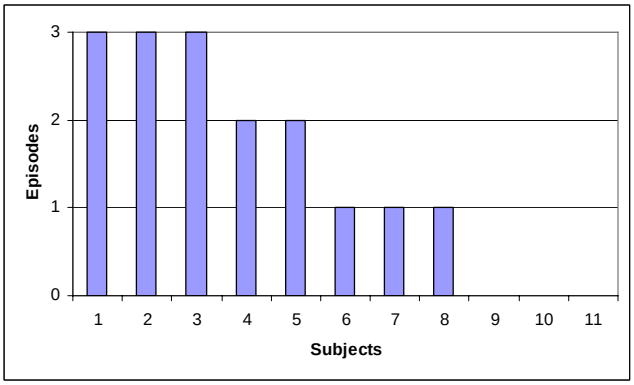

Figure 8. Distribution of observed frustration episodes by subject.

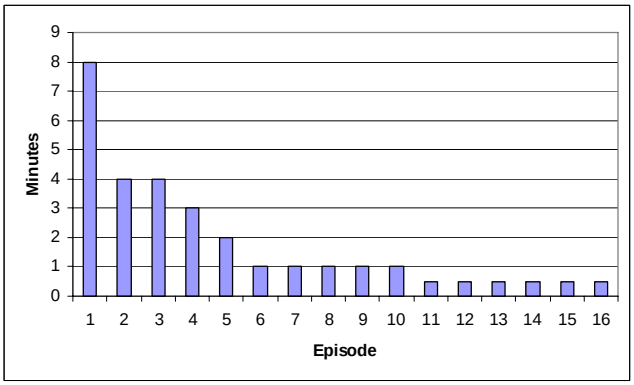

Figure 9. Distribution of length of frustration episodes.

We also observed four additional instances among four of the participants in which they used a work-around but were apparently unaware that there was a better way to do what they wanted. Three of the episodes involved Excel; participants could have saved time and effort in two cases by using copy options and in one case by using a formula. The fourth episode involved Word; the participant could have saved much time and effort by using mail-merge. For these work-around episodes, the mean task time was 20 minutes and the median task time was 10 minutes. We estimate that use of available functions, had the participants known of them, would have

reduced task times by 75 percent. If subjects are unaware of application functions that would let them be more efficient in their work, or otherwise believe that they are doing their work the right way, then they will not report these lost times in surveys or interviews. Our observation of these work-around episodes suggests that survey and interview methodologies likely understate the true impact of frustration with and time lost using computer applications.

5. DISCUSSION

We now to turn to the fundamental question of how well users' self-estimates of solution methods correlate with their actual behaviors as observed in the workplace. Figure 10 shows the distributions of solution methods as estimated by participants in interviews and as observed in participants' work. (Figures 10 through 12 aggregate solutions across categories because further detail was not provided in some of the studies represented, and categories such as "reboot" or "restart" are omitted.) The correlation between the two distributions is not statistically significant. It appears that there is agreement that people almost never use the printed, but that there is little agreement between the self-estimated and observed distributions as to how often people ask someone else, use on-line help, solve the problem by themselves, or give up. As compared with their mean use of solution approaches in observed work, participants tended to underestimate the extent to which they solved the problem themselves and gave up. In contrast, participants tended to overestimate the extent to which they asked someone else and used on-line help. Participants estimated that they used printed manuals in 5.66 percent of cases, but even that minimal number turned out to be an overestimate relative to actual use.

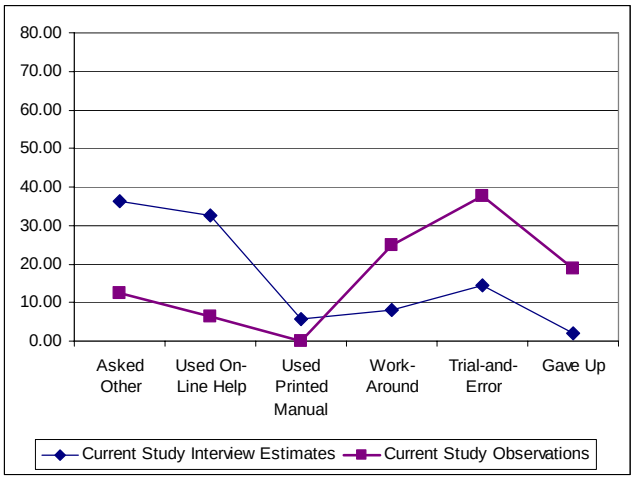

Figure 10. Mean reported or observed solution methods: comparison of interview estimates and direct observation.

These differences between self-estimated and observed solution methods led us to return to the interview data. We extended our analysis by extracting the reported solution methods for each of the frustration episodes mentioned by the participants, aggregating responses across multiple solution attempts, if present, for an episode. The distribution of participants' reported solution attempts for their frustration episode falls roughly mid-way between those for the interview estimates and the observations, as shown in Figure 11. We also conducted an extended analysis of the episode data from the Novick and Ward study [16], and the correlation between these data sets is 0.88.

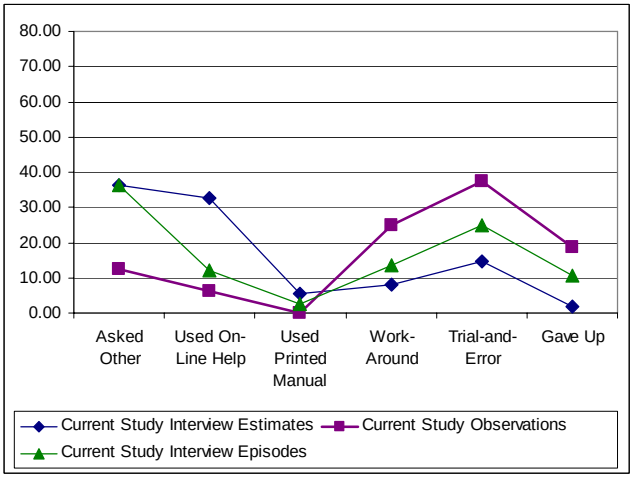

Figure 11. Mean reported or observed solution methods: comparison of interviews estimates, interview episodes, and direct observation.

Taken as a whole, the data suggest that, when interviewed about their approaches to solving problems in using computer applications, people tend to overestimate their recourse to help, even for printed manuals. Given the close correspondence between our interview results and those of Novick and Ward [16], this trend appears reliable. The responses in interviews could reflect a social-desirability bias or an unusually heightened awareness of documentation and help caused by the experimental design. In either case, we conclude that interviews, even asking for specific frustration episodes, (and, for similar reasons, probably surveys as well) are unreliable indicators of people's actual problem-solving approaches for problems with computer applications.

Figure 12 presents the distribution of solution methods for all four studies in which comparable data are obtainable. The current study is represented for interview self-estimates, interview frustration episodes, and for observation. The Novick and Ward study is represented by interview self-estimates and interview frustration episodes. The Mendoza and Novick study is represented by self-reports. And the Ceaparu study is represented by self-reports and observations. The disparities among the distributions likely reflect differences in sampled populations, tasks, and methodology. The one element that all of these disparate results have in common is that users of computer applications rarely use printed documentation. The modestly higher figures in some studies for use of printed documentation likely reflect social desirability bias in the responses. This was probably the case as well for other studies using interview and survey methods (e.g., [18], [11]) not represented in Figure 12.

These results do not mean that interviews are without value. To the contrary, interviews enable researchers to go into greater depth in exploring users' attitudes toward documentation, looking at likes and dislikes, understanding individual frustration episodes, and providing guidance for designers of documentation and help systems. However, interviews appear to be of limited value in determining the overall distribution of users' solution methods.

The comparative value of participative evaluation, as reported by Ceaparu et al. [4] and Mendoza and Novick [12] remains unclear. As indicated in Figure 13, the observations and self-reports obtained

by Ceaparu et al. matched well; the correlation coefficient is 0.87. In contrast, the self-reports from Mendoza and Novick [12] have virtually no correspondence with the self-reports of Ceaparu et al. (correlation coefficient = 0.16) nor with the observation results of the current study (correlation coefficient < 0.01). It may be coincidence, but the correlation coefficient between the Ceaparu self-reports and the observations in the present study is 0.93. In interpreting the correlations between studies, it should be born in mind that we collapsed and omitted some solution categories to make possible the comparisons.

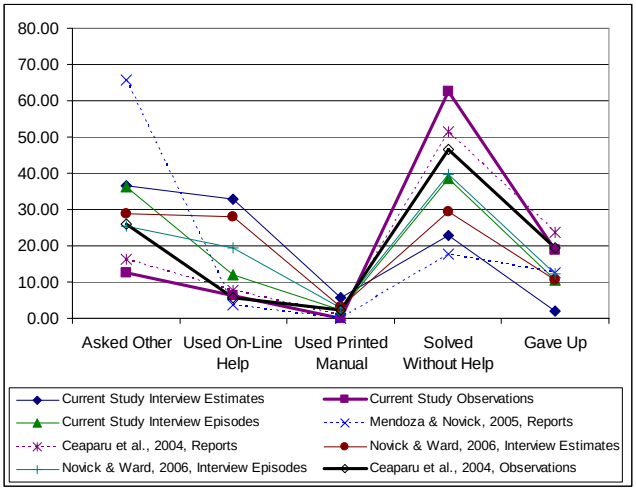

Figure 12. Mean reported or observed solution methods: direct observation (bold lines), interviews (thin lines), and participative evaluation (dotted lines).

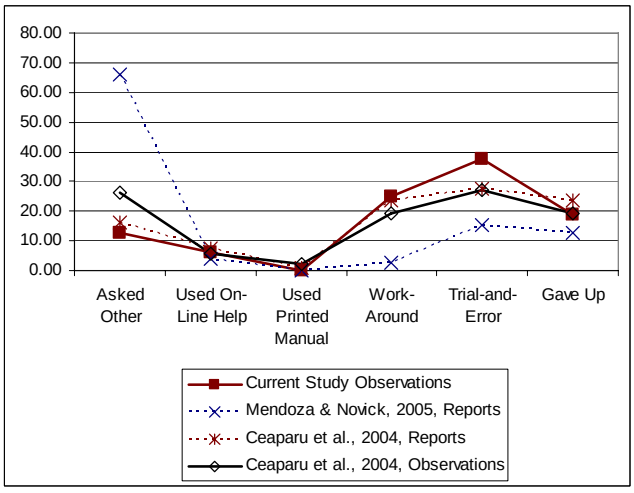

Figure 13. Mean reported or observed solution methods: comparison of direct observation and participative evaluation.

The differences between the two studies using participative evaluation apparently arise from differences in subject populations (university computer science students vs. middle-school teachers) task (diverse tasks vs. a single task), and environment (performing different tasks independently in a computer lab vs. performing the same task with colleagues nearby). Thus we expect that the availability of colleagues working on exactly the same problem accounts in large part for the disparities between (1) "asked other" and (2) the combination of "solved without help" and "gave up." In what may be a depressing perspective on life in the workplace, our observed participants were even less likely than the computer science students to ask someone else, although they were also less likely to give up. Additionally, the "solved without help" category includes work-arounds, and the middle-school teachers may have been able to reduce the number of work-arounds by finding a colleague who already had the solution. In any case, while the correlation between the distribution of solution methods of university computer science students and those of mostly white-collar professionals in the workplace may be coincidental, the degree of correlation nevertheless remains striking.

6. CONCLUSION

Results of our interviews suggest that studies that categorize people's methods of solving problems when using computer applications can present a clearer picture of people's behaviors if broad categories of solution methods are reported in greater detail. Users report that they consult a help desk slightly more often than they ask a friend or colleague and that they use a search engine nearly half as often as they use local on-line help.

Results of our observations suggest that users of computer applications encounter a usability problem on average about once every 75 minutes and typically spend about a minute looking for a solution. But users consume much more time when they are unaware of a direct way of doing something and so end up slogging through a task with ineffective methods; users typically spend 10 minutes on this sort of task, which probably could have been completed in two to three minutes. Survey and interview methodologies thus likely understate the true impact of frustration with and time lost using computer applications. The interview results also suggest that self-estimates of solution methods, and to a lesser extent self-reports of specific solutions, tend to overstate recourse to help.

From these results, we conclude that it remains a major open issue for research as to why users muddle through instead of seeking help. In some cases, users apparently do not know that there is a better way of doing things. For these users of computer applications, their work-around is, so they believe, the solution. In other cases, the users suspect that a better method exists but do not expend the effort to find it. For example, a participant told us that she thought the application probably had a function for summing numbers, but she used a calculator anyway. As they gain experience with a computer application, why do people settle for asymptotic mediocrity? Are there ways of persuading people to seek help more often?

One odd insight into modern work lives emerging from the observations is that it is difficult to determine, from the transcript alone, the participants' business or occupation. Despite the variety of work settings and computer applications, there is a palpable sameness to the participants' activities—filling out forms, sending e-mail messages, and producing or modifying documents.

Our data also suggest that interviews, and probably surveys, provide less reliable views of users' problem-solving behaviors than do participatory evaluation and direct observation. The sheer level of effort involved in observation studies, both in recruiting subjects and in conducting and analyzing the observations, tends to limit the scope of this approach. With all of these methodologies, the choice of subjects and applications may lead to big differences in results.

Our own observation study was subject to a choice-of-subject consideration. Because we were limited to the subset of

interviewed subjects who were willing to then let us observe them at work, it is possible that the observed participants do not fairly represent the larger set of interviewed participants. Likewise, the network of acquaintances through which we recruited subjects may not have led to a sample that fairly represented the general population of people who use computers at work. While we took pains to make clear to potential participants that we were looking at computing in everyday life at work, that we not interested in "computer experts," and that we were evaluating the computer applications rather than the participants, it is possible that fear of embarrassment may have dissuaded people from participating, particularly in the observation phase. The potential discrepancy between interviewed and observed participants could be avoided in future research by studying only participants who agreed at the outset to be observed. The more general problem of the representational validity of the sample remains to be solved; results of such studies can be interpreted in light of their reported distributions of occupations of participants.

7. ACKNOWLEDGMENTS

This work was supported by National Science Foundation Award No. 0080940 and an endowment from AT&T. We thank the study participants, especially those who agreed to be observed, for their cooperation. We also thank the reviewers for their suggestions.

8. REFERENCES

[1] Bessiere, K., Ceaparu, I., Lazar, J., Robinson, J., and Shneiderman, B. (2003). Social and psychological influences on computer user frustration, CS Technical Report 4410, Department of Computer Science. University of Maryland.

[2] Björkenstam, C., and Brundell, S. (2000). Recall effects in retrospective statistical individual Swedish surveys. Technical report, University of Stockholm.

[3] Blackmon, M., Polson, P., Kitajima, M., and Lewis, C. (2002). Cognitive walkthrough for the Web, *Proceedings of the Conference on Human Factors in Computing Systems (CHI 2002)*, Minneapolis, MN, April 20-25, 2002, 463-470.

[4] Ceaparu, I., Lazar, J., Bessiere, K., Robinson, J., and Shneiderman, B. (2004). Determining causes and severity of end-user frustration, *International Journal of Human-Computer Interaction*, 17(3), 333-356.

[5] Desurvire, H. (1994). Faster, cheaper!! Are usability inspection methods as effective as empirical testing? In Nielsen, J. and Mack, R. (eds.), *Usability inspection methods*, John Wiley & Sons, Inc., New York, 173-202.

[6] Galliers, J., Sutcliffe, A., and Minocha, S. (1999). An impact analysis method for safety-critical user interface design, *ACM Transactions on Computer-Human Interaction* 6(4), 341-369.

[7] Grasland, I., Kerbiriou, P., Janse, M., Issarny, V., Sacchetti, D., Talamona, A., Leal, A. (2004). *User experience report of the three demonstrators and the external application of the Ozone framework*, Ozone Project, available at http://www.hitech-projects.com/euprojects/ozone/public_docs/del_oct04/ozone-tho-31aug04-d15d-pk.pdf, accessed May 27, 2007.

[8] Hilbert, D. (1998). A survey of computer-aided techniques for extracting usability information from user interface events, Technical Report UCI-ICS-98-13, Department of Information and Computer Science, University of California at Irvine, March, 1998.

[9] Jeffries, R., Miller, J., Wharton, C., and Uyeda, K. (1991). User interface evaluation in the real world: a comparison of four techniques, *Proceeding of the Conference on Human Factors in Computing System (CHI 91)*, New Orleans, LA, April 27-May 02, 1991, 119-124.

[10] Lazar, J., and Norcio, A. (2000). System and training design for end-user error. In S. Clarke & B. Lehaney (Eds.), *Human-Centered Methods in Information Systems: Current Research and Practice*. Idea Group Publishing, Hershey, PA, 76-90.

[11] Martin, A., Ivory, M., Megraw, R., and Slabonsky, B. (2005). Exploring the persistent problem of user assistance. Technical Report IS-TR-2005-08-01, Information School, University of Washington.

[12] Mendoza, V., and Novick, D. (2005). Usability over time, *Proceedings of SIGDOC 2005*, Coventry, UK, September 21-23, 2005, 151-158.

[13] Mentis, H. M. and Gay, G. K. (2003). User recalled occurrences of usability errors: Implications on the user experience. *Extended Abstracts of the Conference on Human Factors in Computing Systems, Ft. Lauderdale, Fl*, 736-737.

[14] Muller, M., Dayton, T., and Root, R. (1993). Comparing studies that compare usability assessment methods: an unsuccessful search for stable criteria, *INTERACT '93 and CHI '93 Conference Companion on Human Factors in Computing Systems*, Amsterdam, April 24-29, 1993, 185-186.

[15] Nielsen, J. (1992). Finding usability problems through heuristic evaluation, *Proceedings of the Conference on Human Factors in Computing System (CHI 92)*, Monterey, CA, May 3-7, 1992, 373-380.

[16] Novick, D., and Ward, K. (2006). Why don't people read the manual? *Proceedings of SIGDOC 2006*, Myrtle Beach, SC, October 18-20, 2006.

[17] Smart, K., Whiting, M., and De Tienne, K (2001). Assessing the need for printed and online documentation: A study of customer preference and use, *Journal of Business Communication* 38, 285-314.

[18] Smart, K., De Tienne, K., Whiting, M. (1998). Customers' use of documentation: The enduring legacy of print, *Proceedings of SIGDOC 98*, September 24-26, 1998, Quebec, Canada, 23-28.

[19] Silberstein, A. (1989). Recall effects in the U.S. Consumer Expenditure Interview Survey, *Journal of Official Statistics* 5, 125-142.

[20] Wharton, C., Rieman, J., Lewis, C., and Polson, P. (1994). The Cognitive Walkthrough Method: A Practitioner's Guide. In Nielsen, J. and Mack, R. (eds.), *Usability inspection methods*, John Wiley & Sons, Inc., New York, 1994, 105-140.