Help-Based Tutorials

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

Oscar D. Andrade
Computer Sci. Dept.
The University of
Texas at El Paso
El Paso, TX 79968
+1 915-747-8959
odandrade@
miners.utep.edu

Nathaniel Bean
Psychology Dept.
The University of
Texas at El Paso
El Paso, TX 79968
+1 915-747-5551
nhbean@
gmail.com

Edith Elizalde
Computer Sci. Dept.
The University of
Texas at El Paso
El Paso, TX 79968
+1 915-747-5480
edith.elizalde@
gmail.com

ABSTRACT

The consensus of the documentation literature is that users rarely use help, usually preferring to muddle through. To increase use of help, tutorials for novice users could be changed from guided presentations toward using the system's actual help system. To determine whether this approach would increase users' use of help when they encountered problems with an application, we developed an alternative, help-based tutorial introduction to Microsoft Publisher. We compared the behaviors of users introduced to Publisher with the help-based tutorial with the behaviors of users who learned from a traditional tutorial. A balanced study of 22 novice users of Publisher suggests that using a help-based tutorial leads to significantly greater use of help systems when users encounter problems. However, the data also suggest that the increased use of help may not lead to more effective task performance.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *Training, help, and documentation.*

General Terms

Human Factors, Measurement

Keywords

Usability, evaluation, tutorials

1. INTRODUCTION

Thousands of technical writers spend years producing manuals and help systems intended to help people use computer applications more effectively. Despite these enormous efforts, users of computer applications routinely muddle through with trial-and-error methods rather than take advantage of the help available to them. It may be possible to redesign the way in which users learn about computer applications so that they are more familiar with and more likely to use help systems.

Although recent research has helped to quantify low rates of use of help systems, this problem has been of concern for many years, notably as the impetus of the trend toward minimal manuals [4], [5]. But even with wide acceptance of minimalism by the documentation community, use of documentation in general, including help systems, remains surprisingly low.

While some studies have reported relatively higher uses of computer documentation and help systems (e.g., [11], [10]), the evidence on balance is that most users, when faced with problems in a computer application, generally do not use documentation or help systems ([7]; see also [3], [6]). Observations of people in work settings suggest that reluctance to use help is even more widespread than indicated by users' self-reports. While some studies relying on interviews (e.g., [10], [7]) reported that people used online help in roughly 20-35% of cases where they encountered a problem with a computer application, studies based on direct observation and participative evaluation indicated that the people used online help in fewer than 10% of such cases (see [3], [6], [9]). Rather than use help, people tend to use less-effective methods and trial-and-error techniques [9].

Frustrations with computer applications arise in part because people are overwhelmed by the large number of functions these applications typically provide [1]. While people may read a tutorial when first encountering a new application, they still have difficulty finding ways of doing things in the application. In the authors' own experience, having any way of doing something tends to fossilize into the one way of doing that thing. For example, a person who crops images in Microsoft Word using the "Format Picture" dialog box may never learn, unless told by a colleague, that this can be done through direct manipulation via an icon on the "Picture" toolbar. Similarly, people will rely on repetitive cut-and-paste methods when, had they known about it, they could have used a mail-merge function [9]. It appears that people who know of one way to do something may not suspect that there is another, easier way to do the same thing. And even when people know that there must be better way, they sometimes do not believe that they would be able to find and use it [9]. This problem becomes more serious as our increasingly complex applications gain multiple ways of accomplishing a task. The users of the applications too rarely go into the help system to find better ways of doing things.

SIGDOC '08, September 22–24, 2008, Lisbon, Portugal.

How, then, can developers of computer applications guide the users of these applications toward online help? Despite acceptance of minimalism among technical communicators, rates of use of help remain low among users of computers. And the complex, redundant functionality of computer applications engenders reliance on muddling through. If, however, users of computer applications, when first encountering an application, would learn to use the application's help system routinely, then these users might access help more frequently when experiencing frustration or encountering a new problem.

In this paper, we explore the idea that users of a computer application might use online help more often if their introduction to the application provided significant experience in finding application functions through the application's online help system. We present two contrasting approaches to application tutorials—traditional and help-based—and describe an empirical study comparing the effectiveness of the tutorials. We report that a help-based tutorial can increase use of help but does not necessarily lead to more effective task performance. We discuss the study's implications, and conclude by reviewing the study's limitations and open questions.

2. APPROACHES TO TUTORIALS

The traditional tutorial, as exemplified by the tutorial provided for Microsoft Publisher, provides step-by-step guidance for learners of a computer application. The material in this kind of tutorial, although it largely contains the same content found in the help system, stands apart from the help system. An alternative approach, described in this paper, creates the tutorial from building-blocks already present in the help system. This approach minimizes duplicated effort in developing help systems and, more important from the user's perspective, familiarizes the user with the use of the help system as an inherent property of the tutorial.

To illuminate the differences between the traditional and the help-based approaches to tutorials, this section describes a traditional tutorial adapted from the original tutorial for Microsoft Publisher and a help-based tutorial that we developed, which covers the same material.

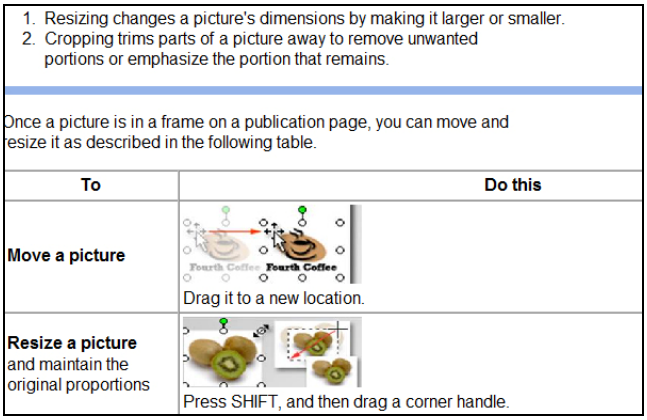
2.1 An Adapted Microsoft Tutorial

An online Microsoft tutorial for Microsoft Publisher served as the basis for both the traditional and help-based tutorials used in this study. Microsoft's original tutorial took nearly 40 minutes to complete and had audio and visual explanations, along with practice and testing modules, on how to use basic features in Publisher such as inserting text and images, editing text and images, and changing image properties. Our adapted version of the traditional tutorial comprised a set of Web pages with step-by-step text explanations with supporting graphics and was completed by participants in a pilot experiment in 15 minutes. The step-by-step explanations had the same wording and order as the original tutorial, but the audio clips and extra modules were removed. Figure 1 depicts an excerpt of the traditional tutorial.

The adapted tutorial and the help-based tutorial included similar instructions on how to access help to find additional assistance, as depicted in Figure 2. The adapted tutorial ended with this material, and the help-based tutorial began with it.

2.2 A Help-Based Tutorial

Our help-based tutorial explained how to use the help system featured in Publisher, as depicted in Figure 2, and gave participants a list of topics for which they should search for assistance. The list of topics was identical to the set of topics covered in both the original Microsoft tutorial and our adapted traditional tutorial; indeed, the topic list came from the quick-reference card provided at the conclusion of Microsoft's tutorial. Participants who used the help-based tutorial were shown a Web page where they read the instructions on how to use help and see topics on which they could search.


1. Resizing changes a picture's dimensions by making it larger or smaller.
2. Cropping trims parts of a picture away to remove unwanted portions or emphasize the portion that remains.

Once a picture is in a frame on a publication page, you can move and resize it as described in the following table.

To	Do this
Move a picture	Drag it to a new location.
Resize a picture and maintain the original proportions	Press SHIFT, and then drag a corner handle.

Figure 1. Excerpt of adapted traditional tutorial.

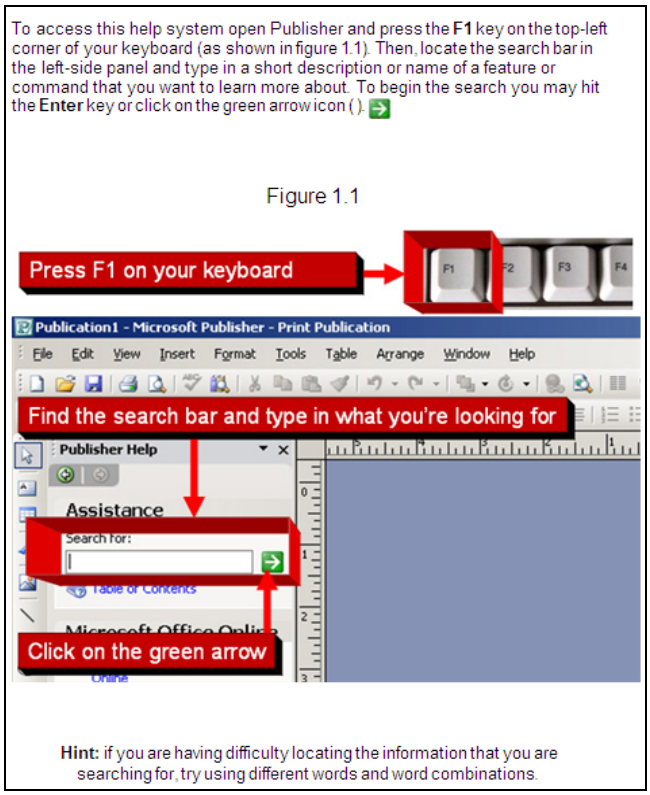


Figure 2. Instructions on accessing help.

Our adaptation resulted in a tutorial enormously shorter than the traditional tutorial. It had the brief introduction on accessing help and the list of topics. The tutorial instructed the participant to "Use

the help system in Publisher to learn about these features." About two thirds of the list of topics is shown in the screen-shot of the help-based tutorial depicted in Figure 3.

While the first Web page of the tutorial is minimal, the help system is not a "minimal manual" in the sense advocated by Carroll [4]. The help-based tutorial includes all of the material in the full traditional tutorial. The difference between the tutorials lies in the involvement demanded of users. The traditional tutorial functions like a book, where short chapters are presented serially. The help-based tutorial has the same chapters, but leads the users to use the help system to reach them. In effect, the users train themselves in use of the help system while nominally learning about the application's substantive features.

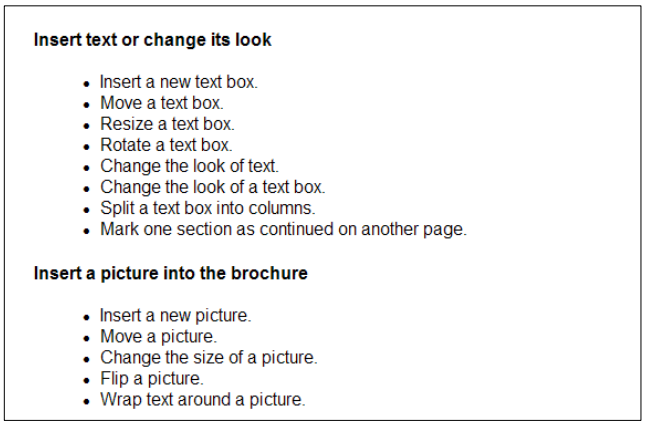


Figure 3. Excerpt of help-based tutorial.

If help-based tutorials train users how to find and use the help system, then these users may be more likely to use an application's help system when they encounter a problem. So in this paper we principally address the question of whether, in realistic settings for use of computer applications, a help-based tutorial will increase use of help systems. And beyond the issue of use of help, we also address the question of whether this increased use of help leads to corresponding improvement in the use of the computer application.

3. METHODOLOGY

This study focuses on determining whether a help-based tutorial would increase the use of help systems and would enable users of computer applications to be more effective in accomplishing tasks. Specifically, we hypothesized that

1. Participants using the help-based tutorial would use help more than participants using the traditional tutorial; and
2. Participants using the help-based tutorial would perform better on novel tasks than participants using the traditional tutorial.

In fact, until the analysis phase of this study, we expected that the results for both hypotheses would be negative. Our experience in observing users of computer applications had suggested to us that changing the tutorial might not have a large enough impact on the participants to change their ingrained behaviors with respect to use of computer applications.

To test these hypotheses, we developed the two tutorials, described in Section 2, to teach people how to use Microsoft Publisher 2003. We conducted a between-subjects experiment comparing the effects of use of these tutorials. In this section, we describe the subjects who participated in the study, outline the experimental design, and briefly describe our post-session survey.

3.1 Subjects

For study subjects, we sought relatively experienced users of Microsoft Office applications who had little or no experience with Microsoft Publisher. We recruited 22 administrative assistants at the University of Texas at El Paso, of whom 21 were professional staff and one was a student employee. Subjects were not compensated for their participation.

The study participants worked in different departments within the university performing clerical work, mainly with Microsoft Office applications. We selected adults between the ages of 22 and 67. All participants were female with an average age of 45.5 and worked as administrative assistants at the University of Texas at El Paso. Participants had, on average, more than 17 years of experience in their profession and spent an average of 7.5 hours a day on the computer. Although we did not select participants on the basis of sex, the prevalence of women among the pool of potential subjects led to having all female participants. All subjects, other than a student employee, had a college education. Most of the subjects had at least 15 years experience as administrative assistants, while the rest had been working for eight years or less. Four subjects indicated that they helped others frequently on using other office applications such as Microsoft Word; twelve helped others with Word less frequently; and six of the subjects did not assist others. Fifteen subjects reported using a computer eight hours a day; one participant reported using it for ten hours a day; and six of the subjects used a computer for six hours or less each day.

Participants self-assessed their general proficiency with computers with a mean 3.18 on a scale from 1 to 5, indicating that they considered themselves to be of average proficiency. Participants self-assessed their specific proficiency with Microsoft Publisher with a mean of 1.64 on a scale from 1 to 5, indicating that they considered themselves to have little or no proficiency with the application.

Even though we designed the tasks to permit all participants to complete the tasks, only differing in the amount of time to completion, the average task completion score was 3.8 out of 5. The participants' assigned experimental condition did not significantly affect the amount of time to task completion (mean 2053.4 seconds, standard deviation 421.0).

3.2 Experimental Design

We used a between-subjects balanced design, assigning subjects randomly to the traditional and help-based tutorial conditions. Each session lasted about an hour and was conducted at the participant's own office, where she used her office computer for the tutorial and tasks. All of the computers ran the Windows XP operating system. Six of the participants had Publisher 2007 and 16 participants had Publisher 2003. We gave each tutorial to every other subject, so that eleven subjects in our study completed the Microsoft tutorial, and the other eleven subjects completed our tutorial.

We asked the subjects to complete a computer-based tutorial that lasted 15 minutes, after which they were given 40 minutes to complete four tasks, where ten minutes were allocated for each task. Subjects were asked to replicate in Publisher every part of a one-page document (the "reference design"), which was handed to them for each task. Subjects were allowed to refer back to their respective tutorials if they needed assistance.

We designed two of the four tasks so that the reference design could be replicated using features covered in both tutorials. The other two tasks depended on features not covered in the tutorials. For each session, the first task had familiar features and was followed by a task with unfamiliar features, which was followed by another task with familiar features, and the last task again had unfamiliar features. The tasks and the features used to create them were

1. *Arrow sign*: change the appearance of a text box, add a picture, rotate a picture, add a background color or color frame to an image (features covered in the tutorial).
2. *Business card*: locate the correct template, change the organization logo, add a decorative frame to a text box (features not covered in the tutorial).
3. *Golden apple advertisement*: divide a textbox into columns, recolor a picture, change how text wraps around a picture (features covered in the tutorial).
4. *Raise petition*: change the color of a picture to grayscale, add a shadow effect to text (features not covered in the tutorial).

Figures 4 and 5 show the reference designs for two of the tasks. We provided the tasks to the participant in the order they were listed above, and we gave the participant ten minutes to complete each task, for a total of forty minutes for the task session. We told the participants that they could use any non-human resource to complete these tasks.

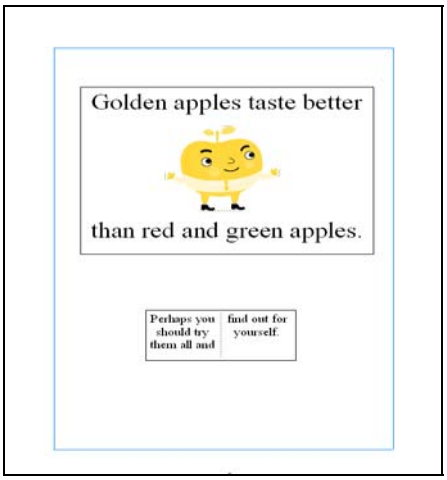


Figure 4. Reference design for Task 3. The apple is a golden color.

3.3 Post-Session Survey

After the tutorial and task sessions, we asked the participants to complete a questionnaire with six questions about their general experience with computers and their attitudes towards learning new software, where for each question each participant had a choice of options ranging from "I strongly agree" to "I strongly disagree".

Ten participants strongly agreed that they enjoyed working with computers; eleven agreed; and one person slightly agreed. Fourteen subjects expressed disagreement with needing other people to help them; five slightly agreed; two agreed; and one subject strongly agreed. Almost all participants reported being comfortable learning new programs, while all participants agreed that learning new programs is a good use of time. Finally, most subjects only slightly agreed that they were better computer users than average, and seventeen strongly agreed that learning about the programs they use at work is a good use of time.

3.4 Data Coding

The audiovisual recordings of the participant's use of Publisher were coded for amount of time (in seconds) spent using the help system and number of times accessing help.

We assessed the participants' task performance by printing the documents they created and having two researchers judge performance relative to the reference design the participants were asked to replicate. We evaluated inter-rater reliability with Cohen's Kappa [2]. Kappa scores for the four tasks ranged from 0.79 to 1.00, and Kappa for overall agreement on the four tasks was 0.89. This suggests that the raters' judgments of task performance were highly reliable.

Figure 5. Reference design for Task 4.

4. RESULTS

Our first hypothesis posited that participants using the help-based tutorial would use help more than participants using the traditional tutorial. As indicated in Table 1, our data suggest that this hypothesis was confirmed. In terms of mean seconds using help, participants using the help-based tutorial used help about eight times more than participants using the traditional tutorial.

	Mean Total Seconds Using Help	Mean Number of Times Help Was Accessed
Help-Based	385.9	5.1
Traditional	46.2	1.2

Table 1. Results for Hypothesis 1.

This difference was significant ($p < 0.05$, one-tailed t-test, unequal variance). This pattern held true for each of the four tasks. Participants in the help-based tutorial also accessed help a greater number of times ($p < 0.01$, one-tailed t-test, unequal variance) than those in the traditional tutorial. These results are particularly striking in light of the small numbers of participants in the two conditions.

Our second hypothesis posited that participants using the help-based tutorial would perform better on the assigned tasks than participants using the traditional tutorial. As indicated in Table 2, the data suggest that this hypothesis is not confirmed. In general, participants who used the help-based tutorial did not perform better on the tasks than participants who used the traditional tutorial.

	Mean Task Performance Rating
Help-Based	3.81
Traditional	3.82

Table 2. Results for Hypothesis 2.

Participants using the help-based tutorial tended to perform slightly better (not significant, two-tailed t-test, equal variance) in Tasks 3 and 4 and slightly worse (not significant, two-tailed t-test, equal variance) in Tasks 1 and 2. Given that Tasks 2 and 4 posed problems not covered in the tutorials, the effects of increased help use should have been greater in these tasks; however the data do not suggest an effect.

5. DISCUSSION

We now discuss related findings and observations, including characteristics of the participants, their use of the tutorials, and possible reasons for the failure to confirm Hypothesis 2.

5.1 Participants' Characteristics

A significant correlation (-0.44, $p < 0.05$) indicated that task performance suffered as participant age increased. Self-reported Publisher proficiency correlated positively (0.52, $p < .05$) with task performance scores. Additionally, individuals who rated themselves as more able computer users performed better on the tasks (Pearson correlation -.53, $p < .05$). These results suggest that participants were, to some degree, aware of their own abilities. Reporting less confidence with learning new computer programs correlated positively (0.45, $p < .05$) with increased use of the help system.

5.2 Participants' Use of the Tutorials

In general, participants who trained with the traditional tutorial read through the bulk of the tutorial and practiced some of the features that were covered in it, only reading the last page, which suggested how they could use help if they needed it, without spending any time in help during the tutorial session. Most subjects finished the tutorial early and did not use the whole 15 minutes available for training. Most subjects practiced using features on a blank page in Publisher as they trained.

Participants who trained with the help-based tutorial read through the tutorial and searched for a few topics and practiced using some features, but tended to practice on fewer features than participants who trained with the traditional tutorial. Participants who used the help-based tutorial tended to explore fewer features and terminated their tutorial session earlier than participants using the traditional tutorial. For example, a participant using the help-based tutorial did not follow directions during the tutorial and was able to accomplish most of the tasks though trial and error. Few subjects spent time practicing features in Publisher because they did not search for many features. Instead, they spent more time reading about the few topics for which they found help.

5.3 Analysis of Results for Hypothesis 2

The negative result with respect to improvement in task performance may reflect differences in the difficulty of the tasks. As indicated in Table 3, the mean task performance for Task 2 was lower than that of the other tasks. This effect was particularly apparent in the help-based tutorial condition.

	Task			
	1	2	3	4
Help-Based	3.59	2.91	4.36	4.36
Traditional	3.95	3.45	3.77	4.09

Table 3. Mean task performance, by condition and task.

As Task 2 required use of functions not covered in the tutorial, the inherent difficulty of this task may have swamped the experimental effect. We did not balance task assignments across conditions, as the Tasks 2 and 4 necessarily had to involve functions not covered in the tutorial. Normalizing the help-based scores by the traditional scores did not clarify the results because the subjects in the help-based condition for Task 2 performed relatively worse than the subjects in the traditional condition. If it were possible to calibrate the tasks in advance of the experiment, it is possible that we could have minimized differences in difficulty between the "covered" and "not-covered" tasks.

Ironically, the positive outcome for Hypothesis 1 may have inadvertently caused the unexpectedly poor task performance of those in the help-based group. Participants in the help-based group spent more time using help and thus took longer to transition to the trial-and-error approach of the participants in the traditional group, who may have felt freer to give up using help. That is, if participants did not find help useful, then time spent in help was time not usefully spent on completing the task.

Consistent with this interpretation, the data for task performance also suggest that Hypothesis 2 was disconfirmed in part because participants used help less as they worked their way through the four tasks. Tables 4 and 5 illustrate our main finding that participants using the help-based tutorial used help longer and more often than participants using the traditional tutorial. Indeed, for the traditional tutorial, these data are consistent with our observation from the literature that users of computer applications in general rarely use help.

	Task			
	1	2	3	4
Help-Based	135.0	104.7	62.1	57.6
Traditional	25.2	12.8	8.2	0.0

Table 4. Mean seconds using help, by condition and task.

	Task			
	1	2	3	4
Help-Based	2.0	1.4	0.9	0.8
Traditional	0.6	0.4	0.2	0.0

Table 5. Mean accesses of help, by condition and task.

Tables 4 and 5 also show that use of help decreased monotonically as a function of task. This result was unexpected, as Tasks 2 and 4 required the use of functions not covered in the

tutorial, which should have led to increased use of help, especially for Task 2, which was the hardest. The trend of decreasing use of help further meant that the participants in neither condition used help more when faced with functions not covered in the tutorials.

If the use of help continued to decrease monotonically, the use would eventually have to asymptote. Indeed, for participants using the traditional tutorial, use of help reached zero seconds and zero accesses by Task 4. This may have been because Task 4 was the easiest of four tasks, given the results in Table 3, but this may also have been because the participants gained proficiency as they made their way through Tasks 1 through 3.

When we reviewed the tasks following the experiment, we discerned that Task 4 was likely easier for participants because, having re-colored the apple in Task 3, they could easily create the effect of graying out the money symbol because they could, and often did, simply re-color the money symbol gray. Applying a shadow to the text proved also to be a relatively less difficult subtask because this function exists in most Microsoft office products. And, as Task 4 was relatively less complex the participants' performances received higher scores for doing less.

Interestingly, though, as shown in the graphical representation of the data in Figures 6 and 7, and despite the relative easiness of Task 4, the rate of decrease of use of help for the participants using the help-based tutorial appears to flatten out markedly for Task 4. A rough extrapolation of the trend for help-based tutorial participants would have otherwise declined to near the zero-level reached by the participants using the traditional tutorial. This suggests that possibly the novel functions required in this task prompted greater use of help than would have otherwise been prompted by the overall trend toward decreasing use of help.

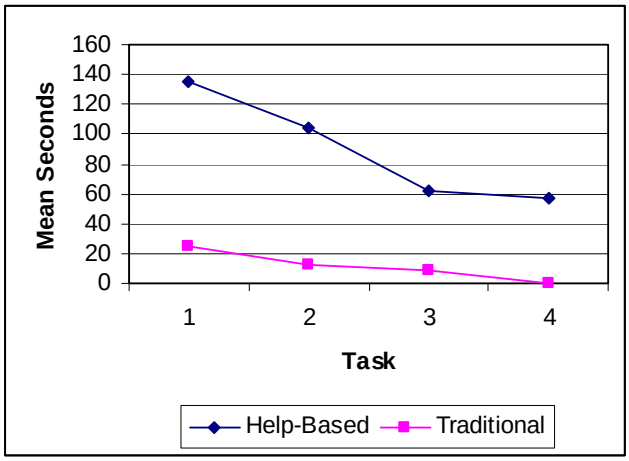


Figure 6. Mean seconds using help, by condition and task.

5.4 Qualitative Findings

Most of the participants in the traditional tutorial condition looked at the traditional tutorial for reference while completing the tasks. Only a few of the participants in the help-based tutorial condition referred back to the tutorial while completing the tasks; rather, they tended to go directly to the help system.

Although we did not design the experiment to stress the participants, some reported that they felt stressed and confused by being asked to complete the tasks. One participant dropped out because she felt that she would not be able to complete the tasks, and another stated as she finished a task, "I failed at this one too, sorry." This comment surprised us because, consistent with best practices for usability tests, we had emphasized that the participants were not to be concerned if they were unable to complete the tasks and that it was probably the system's fault and not theirs. Another participant complained for about five minutes about not being to do the task, not knowing where to start, and so forth. Yet after she started actually working she did a great job relative to the other participants.

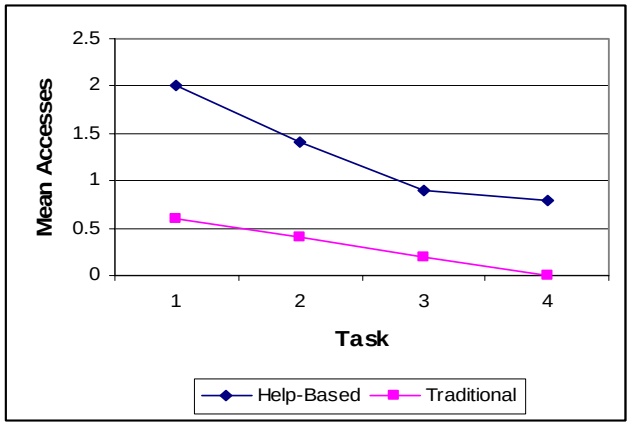


Figure 7. Mean accesses of help, by condition and task.

Consistent with earlier findings [9], we observed that participants used trial-and-error methods for most of their work in trying to complete these tasks.

When participants did use the help system, they often found that it did not actually help them. Participants typically lacked knowledge about the correct terminology to use when searching; sometimes they even looked at the tutorials just to find the right words to use for the help system. One participant even tried using the help system to search for clip art, and another tried using the research system as the help system. The existence of both a local online help and a Web-based online help system led to more confusion. Participants appeared not to understand how the two differed and would often get stuck in the Web-based help system. Moreover, the materials offered by the two help systems are not complementary and often overlapped. One participant, in trying to locate clip art, was led to use the Web-based help system, then led to the online clip art section where she located the correct clip art. If allowed to continue, she would have spent all of her time trying to download that piece of clip art even though it was already available locally on her computer.

When using help, participants spent most of their time searching with different keywords, skimming the results, and sometimes looking into the specific topics. Even within the specific topics participants typically only skimmed the information and often could not correctly use it even when their search had led them to the correct topic. This may occur because participants are unsure if they reached the right topic and know that it would be a great waste of time to read everything presented. It was not uncommon for participants to open up an incorrect help topic, read it for a minute (and even try to use the information to perform a task), and then eventually realize that they were in the wrong topic.

Thus the participants' pattern of skimming, even in the correct help topics, becomes understandable, but compounds their problems in that, when in the correct topic, they skimmed the material and thus were unable to take sufficient advantage of the help available for their task. As a result, they would sometimes abandon both looking for help and working on the task.

Many of the participants approximated the reference design and then gave up, especially with respect to details. Many of the participants' workarounds proved quite effective. Examples of effective workarounds included, among many others:

- Using two text boxes instead of creating columns
- Re-coloring a picture instead of graying it out
- Placing a separate colored box behind a drawing instead of filling the drawing's background
- Drawing a box to put around text instead of coloring the lines of the text box itself.

These workarounds also suggest that participants' understood the reference designs, and thus the tasks of creating them, in ways that differed significantly from the ways in which the reference designs were created.

6. CONCLUSION

Delivering tutorials for computer applications through help-based rather than traditional methods appears to be an effective means of leading users of these applications to rely on greater use of help systems. A between-subjects experiment, using a balanced design with 22 participants, confirmed that a help-based tutorial leads to significantly greater use of help but failed to confirm that this greater use of help led to better task performance. Use of help by participants in both the traditional tutorial and the help-based tutorial conditions tended to decrease their use of help over the four tasks in the experiment. We observed that participants in both conditions also often tended to rely on trial-and-error methods and work-arounds rather than looking for and using more specifically appropriate methods through the help system.

When using the help system, study participants often lacked sufficient knowledge of key words to be able to use the help system effectively. When participants did reach an appropriate topic, they were often unsure that they had found what they needed. As a result, participants tended to skim the help topics and thus sometimes did not learn what they needed even if they reached the right topic.

The complexity of modern computer applications, such the Microsoft Publisher application used in this study, appears to lead to under-use of help in at least two ways. First, users appear to be confused by access to both local and Web-based help, which may cause them to be less eager to try using either path. Second, users may think that the muddle-through method they found by trial and error is, in fact, the right way to solve the problem, which may cause them not to look for help to find a better way. Indeed, given the difficulties we observed in using help, their decision not to use help may be a rational one.

6.1 Limitations

While the study's results for Hypothesis 1 were positive, the study suffered from a number of limitations that might be addressed in future work. These limitations included lack of calibration of the tasks, effects of differences in experimental instruction, and incomplete video records of the tutorial phase of the experiment sessions.

As discussed in Section 5.3, the negative result for Hypothesis 2 may be been in part caused by not having calibrated the four tasks in advance of the experiment. Task 2, as intended, appeared to be more difficult than Task 1. But in the bright, retrospective light of having conducted the experiment, as indicated in Table 3, Task 4 appears to have been as easy or easier than the other three tasks, when it should have been more difficult because it was intended to require knowledge not covered in the tutorial. This lack of calibration did not affect Hypothesis 1, which did not depend on differences among the tasks. But for Hypothesis 2, the lack of calibration possibly affected the results. We note, though, that the remarkable similarity of average task performance across the help-based and traditional tutorial condition (3.81 vs. 3.82 out of 5) suggests that even with calibrated tasks it is unlikely that Hypothesis 2 would have been confirmed.

A second limitation involves the instructions given to the participants, particularly for those using the help-based tutorial. For the participants in the early part of the study, the researchers' instructions for the help-based tutorial were non-specific, such as read the tutorial and complete the task. These participants, on average, used help about three times more than the traditional-tutorial participants. This effect increased for later participants as we were more explicit in the instructions about learning to use the application by using the help system. While our data confirmed Hypothesis 1 with acceptable significance, we expect that both the effect size and level of significance would have been increased had all participants using the help-based tutorial received the more directive version of the instructions.

When we designed the study, we had focused on collecting the data that would confirm (or, as we expected, not confirm) the two hypotheses. These data would be produced through analysis of the task sessions. Thus we recorded the task sessions but did not record the participants' use of the tutorials. As the researchers observed the sessions and began to compare observations, we correspondingly began to appreciate that the participants' behaviors in the tutorial phase were interesting in themselves, even if they did not bear directly on the formal hypotheses. As a result, were able to record the tutorial phase for our final six subjects, three in each condition. The qualitative results in Section 5.4 examining participants' interactions with the tutorial are thus based on recordings of a subset of the participants, augmented by the direct observations of the researchers from the sessions. These results would have been more reliable had we recorded all of the tutorial-phase sessions.

For the reasons discussed, none of these limitations is likely to have led to erroneous confirmation of Hypothesis 1 or erroneous failure to confirm Hypothesis 2. Indeed, the limitations may have led to understating the effect size and significance for Hypothesis 1. The principal effect of the limitations, particularly with respect to the recordings of the tutorial-phase sessions, was to reduce the comprehensiveness of our qualitative analysis.

6.2 Future work

Even though this study involved a formal empirical experiment with well-defined hypotheses, the study's results and, especially, its exploratory aspects, raised a large number of issues that could benefit from additional research:

- Did the participants using the help-based tutorial use (or at least try to use) the help system to find material not covered in the tutorial?
- Is it possible to calibrate tasks? If so, how?
- If use of help induced by help-based tutorials tapers off, are help-based tutorials effective in the longer term?
- We were able to get people to increase their use of help, but this use of help did not appear to provide actual help for the tasks they had. Moreover, as noted above, use of help trended quickly toward zero. So why does the help system not actually help and rapidly lead to discouragement of its users? What are the causes of disappointment with help systems? (Cf., [8]). Are users' implicit calculations of costs versus benefits in seeking help rather than muddling through correct?
- Are the benefits of having access to multiple help systems outweighed by the costs of users' confusion?
- How can help systems aid users in knowing if they have found the right topic for their problem? How else can we get users to avoid skimming and instead learn the material well enough to be proficient?

Some of these questions can be answered, at least in part, through further analysis of the sessions recorded in the study. The recordings, which have high-definition images of the participant's computer screens, can permit detailed analysis of the participants' use of help. Moreover, because the participants' assigned tasks are known, we should be able to assess the participants' strategies for seeking help. Other questions, such as calibration of tasks and longer-term impact of help-based tutorials, will require further collection of empirical data.

7. ACKNOWLEDGMENTS

This work was supported in part by an endowment from AT&T. The authors thank the members of UTEP's Interactive Systems Group for their valuable comments.

8. REFERENCES

[1] Baecker, R., Booth, K., Jovicic, S., McGrenere, J., and Moore, G. (2000). Reducing the gap between what users know and what they need to know. *Proceedings of the ACM 2000 International Conference on Intelligent User Interfaces*, 17-23.

[2] Carletta, J. C. (1996). Assessing agreement on classification tasks: the kappa statistic. *Computational Linguistics*, 22(2), 249-254.

[3] 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.

[4] Carroll, J. M. 1990. *The Nurnberg funnel: Designing minimalist instruction for practical computer skill*. MIT Press.

[5] Carroll, J.M. (Ed.). (1998). *Minimalism beyond the Nurnberg Funnel*. MIT Press.

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

[7] Novick, D., and Ward, K. (2006). Why don't people read the manual? *Proceedings of SIGDOC 2006*, Myrtle Beach, SC, October 18-20, 2006, 11-18.

[8] Novick, D., and Ward, K. (2006). What users say they want in documentation. *Proceedings of SIGDOC 2006*, Myrtle Beach, SC, October 18-20, 2006, 84-91.

[9] Novick, D., Elizalde, E., and Bean, N. (2007). Toward a more accurate view of when and how people seek help with computer applications, *Proceedings of SIGDOC 2007*, El Paso, TX, October 22-24, 2007, 95-102.

[10] 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.

[11] 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.