The role and design of screen images in software documentation

H. van der Meij
Department of Instructional Technology, University of Twente

Abstract  Software documentation for the novice user typically must try to achieve at least three goals: to support basic knowledge and skills development; to prevent or support the handling of mistakes, and to support the joint handling of manual, input device and screen. This paper concentrates on the latter goal. Novice users often experience split-attention problem due to the need to (almost) simultaneously attend to different media. Existing research indicates that split-attention problems can be prevented or reduced by the presence of screen images in the manual. Research is yet unclear about the optimal design of these pictures. This study examines three design styles. Forty-eight novice users received one of the three manual based on these styles. The manuals were an introduction to Windows 95. The users of the most successful manual needed 25% less training time and had a 60% better retention. The most important characteristics of the design style of this manual were its use of full screen images (instead of partial ones) and a two-column lay-out in which the instructions and screen images were presented side-by-side in a left-to-right reading order. The discussion focuses on the tension that exists between theory and practice. Special attention is given to the contributions of a taxonomy of screen images and cognitive load theory.

Keywords: Application software; Cognitive load; Screen images; Training; Tutorial; Visual manuals

Introduction

Early work of Carroll and colleagues has shown that people experience at least three difficulties in trying to get to know a new software program (e.g. Carroll & Mack, 1984; Carroll, 1990). First, one difficulty arises due to the incompatibility of the users’ desire for immediate and meaningful action and the need for learning. Most users are eager to use the program. This orientation towards action is often obstructed by their lack of knowledge as in order to act, they need to know or learn. Second, the users’ tendency to learn by doing leads to numerous mistakes. Mistakes are an inevitable consequence when people try to acquire new cognitive skills through self-instruction. Their desire to reason out and construct understanding...
easily evokes misunderstandings and errors. Finally, difficulties arise due to the need for users to handle simultaneously the program (interface), input device and manual.

Minimalism has outlined a set of design principles and heuristics for handling these three (and other) obstacles (Carroll, 1990; 1998). Empirical research indicates that these minimal manuals have been quite successful in making the training effective and efficient. Compared to conventional (control) manuals, minimal manuals have been found to reduce training time by 25% to 50%, while also significantly and substantially boosting performance afterwards (see the review by Lazonder, 1994). With the overall issue settled, recent studies have focused on optimising particular minimalist techniques for handling a single type of user obstacle. For example, there are studies on finding better ways of supporting the user’s need for meaningful and immediate action (e.g. Gong & Elkerton, 1990; Wiedenbeck et al. 1995) as well as on supporting the handling of mistakes (e.g. Frese et al. 1991; Lazonder & Van der Meij, 1994; 1995). This paper focuses on the split-attention or coordination problems of the user. The nature of coordination problems is briefly described as an introduction to the experimental study that examined three design styles aimed at solving or reducing these problems.

Helping users handle coordination problems

The user of a manual must pay attention to three distinct sources of information: the manual, the screen and an input device. The separate handling of each of these sources generally falls well within what most users can comfortably deal with. It is their joint handling that is taxing (Sweller & Chandler, 1994). Quite a number of mistakes seem to be the result of coordination problems. A typical example is the situation in which users are so busy processing the instructions in the manual that they fail to notice a sign on the screen that the computer cannot process any new input because it is still processing an earlier command. When the user then finally looks up to the screen, the warning has disappeared, leaving the user wondering what happened to the commands just typed in. Novice users probably experience coordination problems quite frequently. Without support for handling these problems, their task execution and learning is at risk.

The best-known and also the most frequently employed means for alleviating coordination problems is presenting screen images. Research indicates that the user can benefit considerably from the presence of these pictures. For example, Van der Meij (1996) found that the users of a manual with screen images completed training significantly faster than did users of a textual manual. The study even reports an effect size of more than one standard deviation. Similarly, Sweller & Chandler (1994) found a shorter training time and better learning outcomes when comparing a manual with screen images with a textual manual.

The study of Gellevij et al. (1999) cautions against an uncritical view on the use of screen images. Their research indicates that the user benefits from these pictures only when their design adequately supports the user’s thoughts and actions. They argue for a careful consideration of both the roles for screen images and their designs (see also Van der Meij & Gellevij, 1998). Following this study, the research on screen images went along two different paths. One study set out to further investigate the roles of screen images (Gellevij et al. submitted). Another study, the one reported here, set out to explore the design issue.

An inventory of the various ways in which screen images are presented in
software documentation shows variations on four important design dimensions: coverage, positioning, size, and cueing (Van der Meij, 1999). Coverage refers to the desktop, window, window element, or object displayed in the screen image. Typically, manuals present full screen images (e.g. the desktop or the application window), partial screen images (e.g. a dialog box or an icon) or a mixture of full and partial screens. Positioning designates the placement of text and screen images vis-à-vis one another. Screen images can be presented so that they are visibly separated from the text (e.g. instruction), or they may be displayed within the text. Size refers to the reduction in size of the window or element in the screen image as compared to actual screen size. Practitioners recommend a constant reduction rate of between 50 and 75% of actual size for all screen images in a manual (Horton, 1991). Cueing refers to the presence or absence of signalling techniques. Signalling techniques are intended to draw the users’ attention to relevant window element(s) or object(s). These may include the use of (coloured) hairlines, circles, call-outs and blurring.

There is an endless set of design possibilities. In the absence of much theoretical or empirical research to guide the design choice(s), this study set out to examine some existing and possibly effective styles. In a sample of 100 manuals two promising examples were discovered: the Instruction – Partial Screen approach (I-PS, see Fig. 1), and the Full Screen – Instruction approach (FS-I, see Fig. 2).

Both design styles are characterised by an abundant use of screen images, presented in a consistent fashion throughout the manual. The two styles have also led to a series of style consistent ‘look-a-like’ manuals for a variety of software programs. Just like the well-known ‘X for dummies’-type manuals, these manuals are commercially produced. Their publisher reports high sales figures and claims high effectiveness (‘the fastest and best way to learn . . .’). Apart from a study by Van der Meij (1996), no scientific studies about their effects have been reported in the literature, however. This study set out to do so. In addition, a third type of visual approach was created namely the Instruction – Full Screen manual (I-FS) displayed in Fig. 3. This approach combines the strong points of the others (Jenné, 1998).

Each approach leads to a different experimental condition in the study. The design differences between the three approaches or styles are summarised in Table 1. Thus, the conditions vary on more than just one design dimension. The study therefore misses some of the elegance of more controlled experimentation. An important advantage of the present set-up is that the study gains validity. It can indicate which of these existing styles works best.

Table 1. A characterisation of the three approaches or styles on the four design dimensions.

<table>
<thead>
<tr>
<th></th>
<th>Instruction - Partial Screen</th>
<th>Full Screen - Instruction</th>
<th>Instruction - Full Screen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coverage</td>
<td>Mainly partial screen images, some full screen images</td>
<td>Full screen images</td>
<td>Full screen images</td>
</tr>
<tr>
<td>Positioning</td>
<td>Separate column design Pictures on right side &amp; Double display of the mouse (occasionally a keyboard)</td>
<td>Separate column design Pictures on left side</td>
<td>Separate column design Pictures on right side &amp; Single display of the mouse (occasionally a keyboard)</td>
</tr>
<tr>
<td>Size</td>
<td>Variable reduction rate</td>
<td>Fixed reduction rate at about 35% of true size</td>
<td>Fixed reduction rate at about 35% of true size</td>
</tr>
<tr>
<td>Cueing</td>
<td>Occasionally hairlines to key elements or objects</td>
<td>Always hairlines to key elements or objects</td>
<td>Always hairlines to key elements or objects</td>
</tr>
</tbody>
</table>

Screen images in software documentation

Fig. 1 A typical section from the Instruction – Partial Screen (I-PS) manual.

Opening a document

After a short moment the WordPad window appears. It is then possible to open your textpaper. You can do this as follows:

1. Click Edit
2. Click Open

Fig. 2 A typical section from the Full Screen – Instruction (FS-I) manual.

Opening a document

Fig. 3 A typical section from the Instruction – Full Screen (I-FS) manual.

The three styles are expected to differ significantly in effectiveness (i.e. training time and learning outcome). The predicted rank order from best to worst manual is: I-PS, FS-I, I-FS. The prediction is based on the following considerations. An important advantage of the Full Screen manuals as compared to the Partial Screen manual is the depicted display of a complete user interface. Full screen images show objects in context and thereby can speed up object location. In addition, the series of pictures in the two manuals convey a sense of continuity and may help users form a mental model of the program (Van der Meij & Gellevij, 1998). The Instruction – Full Screen manual is expected to do better than the Full Screen – Instruction manual mainly because the right-hand column presentation of screens fits the user’s reading direction. In addition, the novice user may benefit from the depicted input device.

Method

Participants
Participants were 48 adult volunteers, 22 men and 26 women, with a mean age of 45.2 years. All participants were inexperienced computer users. Forty-five percent of the participants had no previous experience working with computers and none of the participants had ever worked with Windows 95 (the topic of the manual). Participants were evenly and randomly assigned to the three conditions.

Materials
All manuals were designed as minimal manuals. Each manual gave exactly the same conceptual and procedural information to the user; the contents of the three manuals were identical. In contrast, the typographic designs of the manuals varied considerably. These design differences affected the physical size of the manuals. The two Full Screen manuals consumed 32 pages. The Instruction – Partial screen manual was considerably more bulky, requiring a total of 56 pages.

Training time was recorded automatically via a logging program that also recorded the users key-presses during training. Learning outcome was assessed with a paper-and-pencil test that asked participants to detect, diagnose and correct possible errors. The test contained four questions about trained items and one question about an untrained one. Fig. 4 shows a sample test question.

Procedure
All participants were directed to a computer room in which each person was seated at a computer with a manual. After a brief welcome, participants were told that the study aimed to find what people do and learn in training sessions such as these. They were told that their typing actions would be recorded automatically on the computer. Participants were not informed about the specific interest in manual types and those in the same session always worked with the same type of manual to prevent them from knowing that different types were being used.

Participants answered general questions about their age, sex, educational level and computer experience before being instructed to work on their own and at their own speed. They were to ask the experimenter for help only when they could not solve a problem themselves. Participants were told that the maximum training time of two hours would allow all of them to complete training.

After training and a brief break, the participants were told they were to be tested and invited to take the test immediately or practice some more until they were ready.
to take the test. Some participants, a few in each condition, used this opportunity for further practice (maximally 15 min.). Thereafter, participants completed the learning test during which they were not allowed to access the computer or manual.

**Item 3:**

Align the sentence with the place and date to the right

**Answer:**

![Image of a text editor window with a letter]

**Answer to item 3**

1. Was there a mistake?
   - no
   - Continue with item 4
   - yes
   - Circle the place in the picture where something went wrong
   - Answer the following questions

2. What is the cause of the mistake (why is it wrong?)

3. How can you solve the problem?

The test item presents a goal and an action in pursuit of that goal. Participants must tell whether the intended action fits the goal. If the intended action -the solution- is incorrect, the participant must diagnose it and describe how to progress from the displayed state to the correct solution.

**Fig. 4.** A sample test item.

**Data analysis**

Training time is the time users took to complete their manual. That is, the time before the break. For 13 participants no time score was available. One reason for this loss of data was a weak spot in the logging program which shut down after participants (accidentally) pressed the Esc-key. Participants receive a maximum of two points for each question in the learning outcome test: one for indicating the mistake and one for describing the correct solution. This lead to a maximum score of eight for trained items and a maximum score of two for untrained items.
Checks on the random distribution of participants to the conditions showed no significant effects of personal characteristics none of which was also found to affect training time or learning outcome with the exception of age. Age had a statistically significant effect on training time and, therefore, will be entered as a covariate in the analyses of variance for training time. For learning outcome the findings from univariate analyses of variance (ANOVA)s will be reported.

Results and conclusion

Training time
Table 2 shows the time participants needed to complete training. The results hint at the predicted superiority of the Instruction – Full Screen manual. Participants who worked with the Instruction – Partial Screen and the Full Screen – Instruction manual needed, respectively, 23% and 32% more time to complete training. However, although the difference is considerable, it comes nowhere near a statistically significant level ($F_{2,31} = 0.40, MSE = 131.35, n.s.$).

This absence of a significant effect of manual type on training time may have been due to the presence of a very strong age effect. Age correlated highly with training time (Pearson correlation, $R = 0.59$), accounting for a significant 33% of the variation. Older participants took much longer to complete training than did younger participants. Importantly, the effect of age was the same regardless of the kind of manual employed by the users.

Table 2 shows that participants differed greatly in how long they worked with the manual. Participants’ training time varied between a minimum of 33 mins. and a maximum of 101 mins. Since all manuals required about the same minimum, this suggests that the different design styles mainly affected the maximum amounts of training time. The high variability in training time found in this study is in line with the results of other studies on software documentation (e.g. Carroll et al. 1987; Van der Meij, 1992).

Learning outcomes
Manual type affected the learning that resulted from the training. Participants who had worked with the Instruction – Full Screen manual outperformed the participants working with the Full Screen – Instruction manual and the Instruction – Partial Screen manual with, respectively, 59% and 86% on the trained items (see Table 3). The difference is significant ($F_{2,45} = 4.54, MSE = 3.89, p < 0.05)$, with manual type accounting for 17% of the variance. Scores on the untrained test items were also best for the Instruction – Full Screen manual but this difference was not statistically significant. ANOVA analyses of the learning outcomes of the 35 subjects from which training time was recorded revealed nearly identical scores with those of the total sample of 48 subjects and are therefore not reported.

Table 2. Mean time (in minutes) needed to complete training.

<table>
<thead>
<tr>
<th>Training time (max. 120)</th>
<th>m</th>
<th>SD</th>
<th>range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instruction - Full Screen ($n = 12$)</td>
<td>53</td>
<td>19</td>
<td>33-81</td>
</tr>
<tr>
<td>Full Screen - Instruction ($n = 11$)</td>
<td>70</td>
<td>24</td>
<td>37-101</td>
</tr>
<tr>
<td>Instruction - Partial Screen ($n = 12$)</td>
<td>65</td>
<td>19</td>
<td>35-95</td>
</tr>
</tbody>
</table>

$m = \text{mean}; SD = \text{standard deviation}; n = \text{number of participants}$
In short, the Instruction – Full Screen manual came out as most effective and probably more efficient compared to the other manuals tested. Users who worked with this manual completed training at least 25% faster and showed at least 60% better retention afterwards than users of the other manuals. The finding contradicts the idea that longer training leads to more learning. In this study there was no such trade-off. Instead, the outcome suggests that a shorter training can be better and lead to more learning.

Discussion

Variations in the design of screen images can have a strong impact on the user. Research on positioning and coverage suggests that both are critical factors (Van der Meij, 1998; Gellevij et al., 1999 submitted). For positioning a right-hand placement of screen images seems optimal, whereas for coverage the best design solution seems to be a display of full screens. These are ‘best bets’, based on what research has shown until now. By outlining an agenda for further study, the next paragraphs indicate why research has yet to go a long way before one can draw firm conclusions about the optimal design solutions. The discussion addresses the following issues:

• examining design styles as opposed to a single design variation;
• measuring the quality or usability of a manual;
• examining different user modes or contexts for use;
• adding process measures to the predominantly product-oriented research.

The studies on various kinds of visual designs typically compare design approaches or styles. They examine design variations that differ from each other in more than one way. For example, the pictures in the Full Screen manual in the study of Gellevij et al. (1999) varied not merely on coverage from the pictures in the Part Screen manual. They differed also on positioning, size and cueing. The Partial Screen manual presented the pictures in a left column whereas the Full Screen manual used the right column for their presentation. Variations in size were common in the Partial Screen manual, whereas nearly all the picture in the Full Screen manual were of the same magnitude. Cueing was regularly used in the Full Screen manual. In contrast, no special cueing was used in the Partial Screen manual because the chosen pictures already draw the user’s attention to the main object(s) of interest. Style-oriented research has the advantage of practical relevance. In addition, it can yield suggestions about design differences that matter; it can speed up research by hinting at variations that are likely to make a difference. In view of the many design solutions that are possible, this is not a trivial issue.

Gellevij et al. (submitted) have just completed a function-based study on screen images which forms the complementary part of design-style oriented research. The theoretical basis for the study came from the taxonomy of screen images presented

<table>
<thead>
<tr>
<th>Table 3. Learning outcomes immediately after training.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trained items (max. 8)</td>
</tr>
<tr>
<td>m   SD  range             m   SD  range</td>
</tr>
<tr>
<td>--------------------------------------------------------</td>
</tr>
<tr>
<td>Instruction - Full Screen (n = 16) 4.3 1.9 0-8</td>
</tr>
<tr>
<td>Full Screen – Instruction (n = 16) 2.7 2.1 0-6</td>
</tr>
<tr>
<td>Instruction - Partial Screen (n = 16) 2.3 1.9 0-6</td>
</tr>
</tbody>
</table>

$m = \text{mean; } SD = \text{standard deviation; } n = \text{number of participants.}$

Higher scores mean better learning outcomes.

In short, the Instruction – Full Screen manual came out as most effective and probably more efficient compared to the other manuals tested. Users who worked with this manual completed training at least 25% faster and showed at least 60% better retention afterwards than users of the other manuals. The finding contradicts the idea that longer training leads to more learning. In this study there was no such trade-off. Instead, the outcome suggests that a shorter training can be better and lead to more learning.

Discussion

Variations in the design of screen images can have a strong impact on the user. Research on positioning and coverage suggests that both are critical factors (Van der Meij, 1998; Gellevij et al., 1999 submitted). For positioning a right-hand placement of screen images seems optimal, whereas for coverage the best design solution seems to be a display of full screens. These are ‘best bets’, based on what research has shown until now. By outlining an agenda for further study, the next paragraphs indicate why research has yet to go a long way before one can draw firm conclusions about the optimal design solutions. The discussion addresses the following issues:

• examining design styles as opposed to a single design variation;
• measuring the quality or usability of a manual;
• examining different user modes or contexts for use;
• adding process measures to the predominantly product-oriented research.

The studies on various kinds of visual designs typically compare design approaches or styles. They examine design variations that differ from each other in more than one way. For example, the pictures in the Full Screen manual in the study of Gellevij et al. (1999) varied not merely on coverage from the pictures in the Part Screen manual. They differed also on positioning, size and cueing. The Partial Screen manual presented the pictures in a left column whereas the Full Screen manual used the right column for their presentation. Variations in size were common in the Partial Screen manual, whereas nearly all the picture in the Full Screen manual were of the same magnitude. Cueing was regularly used in the Full Screen manual. In contrast, no special cueing was used in the Partial Screen manual because the chosen pictures already draw the user’s attention to the main object(s) of interest. Style-oriented research has the advantage of practical relevance. In addition, it can yield suggestions about design differences that matter; it can speed up research by hinting at variations that are likely to make a difference. In view of the many design solutions that are possible, this is not a trivial issue.

Gellevij et al. (submitted) have just completed a function-based study on screen images which forms the complementary part of design-style oriented research. The theoretical basis for the study came from the taxonomy of screen images presented
by Van der Meij & Gellervij (1998). More specifically, the study examined whether screen images would facilitate the processes of object identification and location, screen state verification, and the creation of a mental model. The study contrasted a visual manual with a textual manual (the control condition). The designs of the screen images in the visual manual depended completely on their presumed role(s) for the user. This led to the development of a unique design style for each of the three main roles of screen images. The findings indicated that screen images had a positive and significant impact on object identification and location, and on mental model development.

Style-oriented research cannot prove whether a single design variation has a significant impact on the user. For drawing such a causal conclusion it is necessary to conduct a more controlled experiment. A candidate for such research on screen images could, for instance, be a study that capitalises on the special qualities of the Instruction – Full Screen approach. The critical design differences between two Full Screen manuals, namely positioning and the presence of an input device, should (and can) be manipulated in a 2 by 2 set-up so that these variations are fully under the experimenter’s control. The research questions would then be: ‘Is there a significant advantage for positioning full screen images after the instructions (i.e. in a separate right-hand column) rather than before?’; ‘Are users helped by the display of input devices?’ and ‘Is there an interaction of positioning and displays of an input device?’

Another issue for debate concerns the choice of the right measures of success. Put into more general terms, the question here is how one can best measure the quality of a document. This is, of course, a heavily debated topic (see Redish, 1995). The various measures can be grouped into text-focused, expert-judgement focused, and user-focused methods (Schriver, 1989). The measures in text-focused methods include readability formulas, writer guidelines and checklists. Expert-focused measures are reviews by peers (other technical writers), by technical or subject matter experts, and by an editor. A weakness of both these methods is that the expert’s input is indirect at best. Thus, they give only second-hand evidence of the usefulness or usability of a document. Because usability is the primary concern for the presentation of screen images in manuals, text-focused and expert-judgment methods are not optimally suited for the present research effort.

User-focused methods give a better indication of the usability of a document because they rely on feedback from the intended audience. They indicate how a user responds to a document. There are two main types of user-focused methods: concurrent testing and retrospective testing. Concurrent testing registers the user’s behaviour in real-time. Data is gathered while the user processes a document. The indices include observations, time registration, eye-movement registration and verbal protocols. In retrospective testing feedback is elicited after the user has finished reading and using a document. Quality indices may come from tests taken after training (e.g. learning outcomes), but also from surveys, interviews, and reader response cards.

In research, the prevailing quality measure gathered in concurrent testing is training time. Training time has two important advantages over other measures in concurrent testing. The first is simplicity. Not only is it easy to record training time, it can also be done in an unobtrusive manner. Unlike, for instance, think-aloud protocols, the user is not asked to do anything unusual and may remain completely unaware of the fact that time data are being gathered. Another advantage is that,
more than for other indices of quality, it is an argument that appeals to decision makers. Time measures can be used to calculate the financial effects of opting for a particular design. Good documentation can reduce training time/cost of the client.

Learning outcomes prevail in research using retrospective testing. The measure indicates whether the user has acquired the knowledge and skill after training. Although there is a fear that reduced training time (a correlate of ‘time on task’) leads to less learning, the existing research on documentation, most notably that on minimalism, indicates that such fears are unwarranted. Reduced training times have never led to lower learning outcomes than longer training times. On the contrary, shorter training has often been found to be associated with statistically significant better learning outcomes (e.g. Carroll et al., 1987; Van der Meij, 1992; Lazonder & Van der Meij, 1995; Gelluvij et al., 1999). This finding underscores the conclusion of Redish (1995) that various measures of document quality tend to go together.

The measures of training time and learning outcomes have some weaknesses that call for some caution. Training time is not a perfect measure of document quality because the researcher has limited control over some of the total training time of the user. The problem has to do with the time users spend exploring the program. During training, users often ‘fiddle around’. Indeed, these explorations may even be triggered by suggestions in the manual and by some form of conceptual support. What the documentation does not do, and cannot do, is provide detailed support for these explorations. The net effect is that even when the document intends to influence user explorations — as it should according to minimalism — it still cannot control all of the user’s training time spend during these explorations. Separating training time for guided sections and exploratory sections (see Gelluvij et al., 1999) only partly solves this problem.

One should also be cautious about the measures of learning outcomes used in the existing research. The problem here concerns the existence of a large set of measures and limited user testing time. Because learning outcome can be measured in various ways and users cannot be bothered with a complete array of tests, researchers must make a selection of what to measure and how to measure it. An important choice is that of testing with or without the user actually handling the software. The user can be asked to complete the test while working on the computer, or by answering the questions on paper. Another choice concerns the presence of support. The conditions for completing the test may vary from no support at all to a situation in which the user can consult the training manual and has access to online help. A third decision has to do with the choice of test items. These items may measure knowledge or skill, and trained items or untrained (transfer) items. The present study tried to substantiate the prediction that screen images can help develop a mental model. For this reason, the test questions involved some form of problem solving. By asking the user to react only on paper, the experimenter has more control over the tasks of the user. In addition, the procedure facilitates data analyses which otherwise depend on recorded logs. To maintain realism, the test depicted screen states.

Research has yet to put more effort into discovering the impact of various user modes on the roles and designs of screen images. The experimental work on screen images has concentrated on users in a learning mode. With one exception, all the users in the studies were novices who were invited to ‘work in their own way and at their own pace.’ The users were given a tutorial to help them become acquainted with the basics of a new program. In addition, they volunteered to participate in the
study and were willing to spend one or more hours learning a new product.

Such an experimental situation differs considerably from one in which users are focused on achieving real tasks. In such a doing mode, users turn to their documentation with different needs. Reference documentation is designed especially to satisfy these needs. As a result, reference manuals differ considerably from the tutorials used in the training studies. Typically, these manuals organise the subject matter differently, and they have more breadth and depth of presentation of the (new) product.

Only one study has examined the impact of screen images on users in a ‘doing mode’ (Van der Meij, 1998). This study was conducted in a large chemical company that had recently changed to a new database program. The users were asked to perform some of their daily tasks using one of two manuals, a visual or a textual manual, for support. The findings clearly showed that the screen images in the visual manual were helpful. Users of the visual manual completed their tasks much faster than did users of the textual manual. A speed gain of 35% (a difference of more than one standard deviation) was found for screen image supported task execution.

The finding of a positive effect of screen images on users in a ‘doing mode’ is in line with the findings for users in a learning mode. But more research is needed to find out how screen images impact on the two user modes. It is still an open question as to whether each mode requires its own design, or whether one style might satisfy both.

Research should attend more to the actual use of screen images. The majority of studies are heavily product-oriented. Therefore, very little is known about how users process the screen images presented in a manual. Observations of use in practice might indicate, for example, when and how often users ignore screen images, whether partial screen images invite a more detailed or active processing of the user interface than do full screen images, and how, after an error, users react to a mismatch between the screen image and the user interface. Such insights, along with the proposed taxonomy of screen images (Van der Meij & Gellevij, 1998), can lead to a sound basis for creating optimal designs.

In addition, observation studies can contribute to finding optimal conditions for use of the screen images. A prime theory one might consider in this respect is cognitive load (Sweller et al., 1998). The main tenet of this theory is that people have a limited capacity working memory and thus that one should prevent or reduce short-term memory overload. A prototypical moment at which memory overload is likely to occur is during split-attention or coordination. Short-term memory is easily taxed when people must deal with problems that involve the simultaneous handling of different sources of information (e.g. paper and screen), or different means of information (e.g. picture and text). In other words, cognitive load theory attends to the same situation that led to the design of the present study.

While there are thus good reasons to attend to the insights from cognitive load theory, two cautions are in order. First, one should bear in mind that cognitive load theory relates to the design issue only indirectly. It is a psychological theory that primarily predicts when coordination obstacles are likely to occur. Although it has implications for design, it is not a design theory. Second, not all of the designs studied in the experiments on cognitive load satisfy with regard to usability (Carroll & Van der Meij, 1998). For example, in several studies Sweller & Chandler (1994) used manuals that were designed as study guides rather than as manuals supporting
users to practice with the program. Under conditions in which users are free to do as they please, these manuals probably do not work well because they fail to address the important obstacle of the user’s need for immediate and meaningful action.

The outlined agenda indicates that there is still a lot of work to be done. Practitioners, of course, cannot wait for all this research to take place. They are facing their design questions now, and they want an answer to these questions now. Practitioners are usually not insensitive to the needs of researchers, but their primary need typically makes them ask for ‘best bets’ about design styles. The present study is a case in point, as it was stimulated by the creator of the Instruction – Partial Screen (I-PS) design style, Addo Stuur, who invited Marijn Jenné to analyse his work in detail. During that effort, he also helped Marijn and me in creating the Instruction – Full Screen design style (I-FS) that was found to be a superior solution in the present study. Moreover, after the work, Addo Stuur decided to adopt this style for his new documentation efforts. The first product has recently appeared on the open market in the form of a manual introducing Windows 98 to seniors (55+ users). It has become an instant success with over 35,000 copies sold in less than six months. Surely then, the current study is a good example of how theory and practice can join forces, illustrating how the two can inform and influence each other.

Acknowledgements

The experiment reported in this paper was designed and executed in cooperation with Marijn Jenné. I want to thank Addo Stuur for offering the context for this study and for his assistance. I also want to thank Mark Gellevij for his detailed and valuable comments on a draft of the paper. The financial support from the American Society for Technical Communication (STC) for conducting the study is gratefully acknowledged.

References


