Screen Captures to Support Switching Attention

—MARK GELLEVIJ
AND HANS VAN DER MEIJ

Abstract*—*This study set out to validate the supportive role of screen captures for switching attention. Forty-two participants learned how to work with Microsoft Excel with a paper manual. There were three types of manuals: a textual manual, a visual manual with full-screen captures, and a visual manual with a mixture of partial- and full-screen captures. The findings show that participants in all conditions looked up from the manual to the screen on about 97% of the cases in which such a switch was called for. Rank order analyses showed that users of the visual manuals switched attention significantly more often than did users of the textual manual. No differences were found between conditions on learning effects and training time.

Index Terms*—*Documentation, graphics, screen captures, usability, visualizations.

Manuscript received January 10, 2002; revised March 7, 2002.
The authors are with the Faculty of Educational Science and Technology, Department of Instructional Technology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands (email: M.R.M.Gellevij@edte.utwente.nl).
IEEE PII S 0361-1434(02)04973-1.

Users who consult a manual to learn how to use a computer program run the risk of falling into the nose-in-the-book syndrome [1]. That is, they may keep their attention too focused on processing the manual at the expense of paying little attention to what happens on the computer screen. In contrast to this first problem, graphic interfaces may be so attractive and present so much information that users start exploring the system without following the manual's instructions. To remedy these problems, a manual can stimulate users to look back and forth from the manual to the screen regularly. This study set out to examine the function of screen captures that prompt such switching of attention.

Screen captures seem well suited to support switching behavior. They simplify the process of applying the information presented in the manual. In a purely textual manual, the user needs to put in cognitive effort in finding the right match between the textual description and the corresponding pictorial representation on the screen. Screen captures circumvent this problem. With screen captures, the user does not need to connect divergent modalities; instead, the user must match identical representation modes. At the same time, screen captures simplify the process of moving from the computer screen to the right place in the manual. In other words, screen captures offer better support for re-entry than do written statements because they stand out more on the page. They also facilitate access into the manual because they are easier to perceive.

To date, No research has shown that users benefit from screen captures for switching behavior. There are, however, other features of a manual that may stimulate attention switching. Research on minimal manuals has shown that users sometimes carry out as much as 90% of the instructions [2], which shows a great deal of compliance with the action-oriented focus of this manual design. The frequency of

attention switching is likely to be high as well because minimal manuals, along with many recent manuals, tend to get users more easily into what is perhaps best characterized as a "switching mode." When instructions to act prevail over the presence of conceptual information, the manual is more likely to facilitate the development of a switching habit. Still, even in such manuals, screen captures may strengthen switching attention behavior, although the potential benefits then presumably are relatively small.

The goal of this study is to examine the influence of screen captures on switching behavior in an action-oriented manual. The effects of two types of visual manuals and a textual (control) manual were studied. One visual manual presented full screen captures (Visual-Full) and the other contained a mixture of partial and full screen captures (Visual-Part&Full). Figs. 1 and 2 illustrate the two manual types.

The difference between the two visual manuals lies in the representation of the object (i.e., the equals sign) that the user must act upon. In the Visual-Full manual, the user must examine both the screen capture and computer screen to find the object (=). In the Visual-Part&Full manual, the object (=) is specifically shown in the partial screen capture. Finding the object becomes easier, and the Visual-Part&Full manual becomes more self-contained.

When considering the advantages and drawbacks of the two visual variants on the support they can give on switching attention, it is difficult to anticipate their effects. There are three aspects of the screen-capture designs that determine such effects: readability of the screen captures, self-containedness of the manual, and redundancy of screen captures.

Fig. 1. Example of instructions from the Visual-Full manual type.

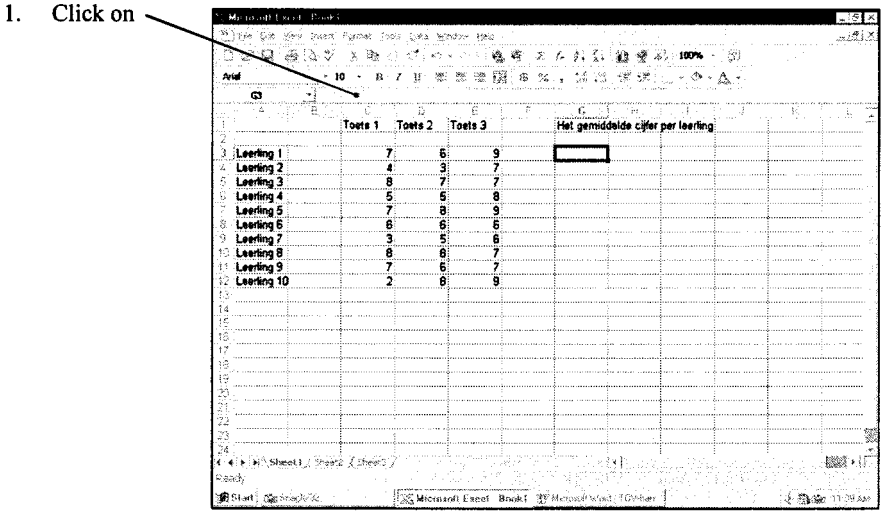
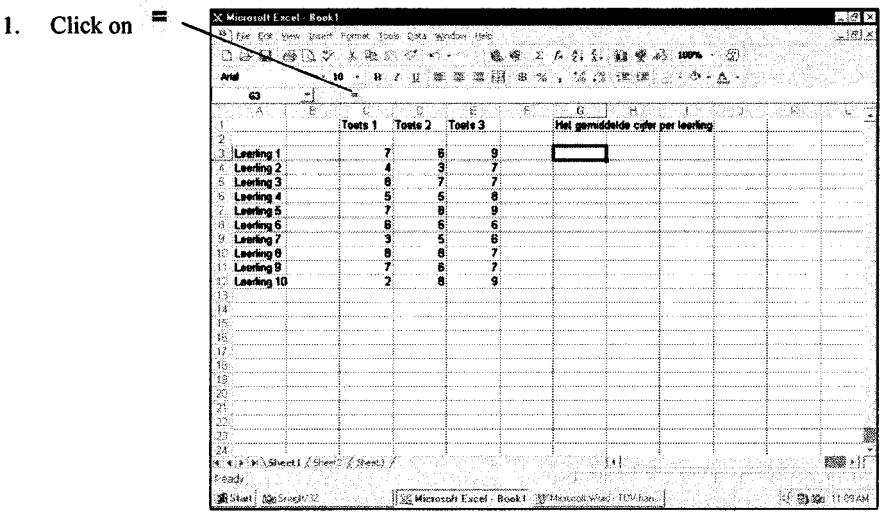


Fig. 2. Example of instructions from Visual-Part&Full manual type.



Regarding the READABILITY of the screen captures, the Visual-Part&Full manual seems the best designed to support switching attention. After all, the object that needs to be clicked on is presented in a readable format and connected to the full screen capture, which makes finding that object on the computer screen relatively easy. Easier than when this partial screen capture is omitted, which is the case in the Visual-Full manual. Here, only the "hard to read" full screen is presented, which may make it more difficult to find the object. It can be expected that the ease with which the relevant object on the screen can be found promotes switching behavior. This way, the Visual-Part&Full manual seems the best design for switching attention.

Regarding SELF-CONTAINEDNESS of a manual, the Visual-Full manual seems the best design to support switching attention. The action step, connected to the full screen capture that is difficult to read, forces the user to look up to screen to find the relevant object. In contrast, because of the combination of the partial and full screen captures, in the Visual-Part&Full manual it is possible to read the instructions without carrying them out. This self-contained Visual-Part&Full manual most probably does not stimulate switching as well as a manual that is less self-contained.

Regarding REDUNDANCY of screen captures, the Visual-Full manual seems the best design to support switching attention. The partial screen capture in combination with the full one in the Visual-Part&Full manual present the same information twice. Both need to be processed in working memory, which may increase cognitive load [3]. Because of that increase, switching behavior may be hindered.

The Visual-Full manual seems to be the best screen capture design for switching attention on two out of the three aspects discussed above. To predict that this design will, therefore, be the better of the two is rather simplistic, as it is hard to envision whether all three aspects have an equal impact on the potential effects. Therefore, no prediction is given about which of the two visual manuals best supports switching.

A textual manual was used in the control condition. The visual manuals were expected to stimulate users to switch attention more often than the textual manual because the screen captures used in the visual manuals simplify the processes of looking up from the manual to the screen and from the screen back into the manual, as argued earlier.

In addition to switching behavior, we also examined training time and learning effects on a post test. Previous research has shown that visual manuals and textual manuals can have very different effects on these variables. However, compared to those studies, the training task in the present study is relatively short, which may make it difficult to find such differences.

METHOD

Participants Forty-two students (6 males and 36 females) from the Faculty of Educational Science and Technology from the University of Twente participated in the study. The computer knowledge and skills of this group, measured by self-grading on a scale from 1 to 10, was moderate to high with a mean score of 6.90 (*SD* = 0.88). Participants had no experience with Microsoft Excel, the topic of the manuals. They participated on a voluntary basis.

Materials

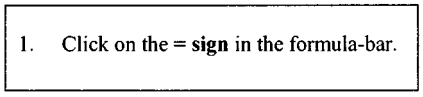
Manuals: The designs of the manuals were based on the findings from previous studies on screen captures. A Guided Tour design [4] was used as the primary basis for the two visual manuals. Gellevij, Van der Meij, De Jong, and Pieters [5] have shown that this manual which displays only full screen captures, leads to more learning compared to a manual that uses only partial screen captures. In the Guided Tour approach, written instructions are presented on the right-hand side of the page with the full-screen capture presented on the left. Van der Meij [6] found that learning improves with a reversed order. Presenting the instructions on the left hand side of the page and the screen captures on the right preserves the reading direction of western audiences and thus better supports the processing of the information in the manual.

The Visual-Full manual is presented in Fig. 1; the Visual-Part&Full manual is shown in Fig. 2. In both manuals, the instruction is connected with the full-screen capture by a thin hairline. More importantly, the written instruction is incomplete without the picture. The user must attend to the screen capture to find the object that must be acted on.

The third manual contained no screen captures. To compensate for the absence of visual cues for identifying and locating screen objects, such information was added to the verbalized instructions in this textual manual. That is, objects that would be depicted in the instructions in the Visual-Part&Full manual were described or typed, along with information about their location on the screen (see Fig. 3.).

The manuals, written in Dutch, consisted of five chapters covering the topics: starting MSExcel, creating a datasheet, changing a datasheet, creating a graph, and

Fig. 3. Example of instructions from the Textual manual type.



closing MSExcel. Apart from the presence of the screen captures, the three manuals were identical in content as well as layout.

Questionnaires and Tests: Participants were asked to think aloud while carrying out the experimental task. This way the observers had two cues for observing switching behavior, namely, the (sometimes slight) movement of the head when users looked up from the manual to the screen and the corresponding thoughts of the user. Switching attention was recorded by jotting down the number of times a user looked up from the manual to the computer screen. These notes were recorded on an observation form which contained all text fragments of the manual. Each instruction (action step) in this form was connected to a check box in which the observer noted the number of times the participant switched attention for that instruction. The observers were also instructed to mark those moments in which they failed to observe switching behavior (e.g., because it could not be ascertained with full certainty that the user switched attention). The two observers did a pilot observation in which they both observed two participants to reach agreement about the use of the observation instrument. In this pilot, interrater reliability (Cohen's kappa) was 0.58, which is considered acceptable [7], [8].

Some of the personal characteristics of the participants were measured with a short questionnaire about gender (male or female), age (in years), touch-typing skill (yes or no), and computer experience (based of self-grading on a scale from 0 (no skill at all) to 10 (extremely well-skilled)).

Training time was measured in seconds by the observer with the aid of a stopwatch. After training, participants received a written test which consisted of five questions. Two questions involved trained tasks and three questions asked about untrained tasks. Trained tasks are tasks that are the same as practiced with the manual. An example of a trained task is: "What happens when you click on the √-icon?." Transfer tasks or untrained tasks test whether participants can apply their recently acquired knowledge and skill to (slightly) new aspects of the software. An example of an untrained task is: "What information is needed before one can use the  -icons?."

Procedure The experiment was held in two adjacent private rooms with the observer seated in one of these rooms and the participant in the other. Participants were told that the goal of the study was to investigate how people act when they learn to work with a computer program using a paper manual as their only form of support. Participants were seated behind a computer and asked to think aloud during training. The observer was seated in the adjacent room which gave an unobtrusive view into the other room through a one-way mirror. During training, the observer registered the participant's actions. The training session took a maximum of 40 minutes. After a short break, the participant was asked to complete the test. Participants were not informed about this test beforehand. After the test, they were informed of the goal and design of the experiment.

Coding and Scoring

Number of Participants in the Analyses: The observation form together with the think-aloud procedure proved to be a workable format for recording switching behavior. Even so, the observers were not able to score the user's reactions to all instructions. This was mainly due to the speed with which some participants processed some of the instructions. The manuals included 45 instructions (action steps). Therefore, there were also 45 moments in which the information in the manual intended a switch between manual and computer screen. For five participants more than 10% of observations for these expected switching moments were missing. These participants were removed from the dataset, yielding a total number of 37 participants in the statistical analyses.

Switching Attention: The observation data were examined in three ways. To assess the relative number of switches the behavior for each instruction was scored as a switch (score 1) or a nonswitch (score 0), regardless of the number of times the user switched attention for that particular instruction. The relative number of switching acts is simply the user's actual score divided by the maximum number of instructions. The theoretical maximum of the latter is 45 given the 45 instructions in each manual. When there are missing values, this maximum is adapted. For example, if the observer had failed to register two instructions of a participant, the actual score for this person would be divided by 43.

Participants could switch more than once for an instruction, of course. Repeated switches may occur, for example, when the user rereads an instruction or wants to be absolutely sure that the screen capture and the actual screen are the same. Such repeated switching acts are taken into account by calculating the absolute number of switches. This measure is simply the mean sum of all switching acts of a participant. For this measure, the maximum score can lie considerably above 45 switches.

We also calculated a third measure of switching. The measure is the mean rank order of switches. In this measure, the three conditions are given a rank score for each of the 45 instructions on the basis of the mean number of switches, where a high mean amount of switches leads to a high rank. For example, in the three conditions,

the mean number of switches for the first instruction were 1.25 for the Visual-Full, 1.18 for the Visual-Part&Full, and 1.36 for the Textual condition. This led to the following rank order for the first instruction: 1 for the Visual-Part&Full manual, 2 for the Visual-Full manual, and 3 for the Textual manual.

Learning Effects: The two trained test items and three untrained test items were divided into distinct test elements. For each correct answer to such an element a participant received one point. As both trained test items consisted of two elements, this led to a maximum score of 4 for the trained tasks. The three untrained task items consisted of 7, 2, and 1 element(s), leading to a maximum score of 10 for the untrained tasks.

RESULTS

Check for Randomization For the covariates computer experience and age, Anova's were used to check the random distribution of participants over conditions. For the same reason, Chi-square tests were conducted for the covariates gender and touch-typing skill. The findings show that there were no statistically significant differences between conditions on personal characteristics except for age (see also Table I). Tukey tests revealed that participants in the Visual-Full condition were younger than the participants in the Textual condition (F (1, 36) = 3.962, p = 0.028). Although this finding shows that the participants were not randomly distributed over conditions based on their age (see Table I), there is no theoretical assumption that foresees a disturbing influence of this variable. Age is therefore not treated as a covariate in the analyses. (The results on training time, learning effects, and switching attention show a similar pattern with and without age as a covariate.)

Training Time and Learning Effects There were no differences between conditions on training time and on learning effects (see Table II). Most of the participants spent less than half an hour on learning how to work with the program. They scored 87% correct on trained tasks and 48% correct on untrained tasks.

Switching Attention The behavior of participants on switching attention was tested in three ways: in a relative, absolute, and rank order manner. Table II shows the mean scores for each condition on the percentage of switches (relative) and on the amount of switches (absolute). ANOVAs showed no statistically significant differences between conditions for these measures. In about 97% of the possible moments in which a switch was called for and expected, users did look up from the manual toward the screen at least once. Furthermore, the mean absolute number of switches lies between 58 and 60 times indicating that the average users switches about 1.3 times for an instruction. Here too, no differences between conditions were found. The reason that this average is higher than one switch lies in the fact that some action steps in the manual contained more than one task. The peaks for the amount of switches for the

TABLE I
THE PARTICIPANTS' AGE (WITH STANDARD DEVIATION), GENDER, AND ABILITY TO TOUCH-TYPE BY MANUAL TYPE USED IN EXPERIMENTAL CONDITIONS

Manual		Gender		Touch-type Ability	
	Age (in years)	Female	Male	Yes	No
Visual-Full	20.50 (1.24)	11	1	5	7
Visual-Part & Full	20.72 (2.20)	9	2	4	7
Textual	22.21 (1.58)	12	2	3	11

TABLE II
MEANS (WITH STANDARD DEVIATIONS) ON TRAINING TIME, LEARNING EFFECTS, AND SWITCHING ATTENTION BY MANUAL TYPE USED IN THE EXPERIMENTAL CONDITIONS

Manual		Learning Effects		Switching Attention	
	Training Time (in seconds)	Trained Tasks (max=4)	Untrained Tasks (max=10)	Percentage of Switches	Absolute No. of Switches
Visual-Full (n=12)	1645 (317)	3.17 (1.11)	4.58 (1.56)	97.14 (2.60)	59.17 (6.93)
Visual-Part & Full (n=11)	1592 (181)	3.55 (0.69)	5.55 (2.11)	97.06 (2.58)	59.64 (11.12)
Textual (n=14)	1584 (266)	3.71 (0.47)	4.29 (2.23)	96.89 (2.76)	57.57 (6.39)

actions steps 4, 7, 19, and 26 in Fig. 4 can be explained by this. These steps asked, for example, to *Type in cell A4 to A12, student 2 to student 10*, or to *Change the formulas in cells G4 to G12 yourself*. This is in contrast to the majority of action steps that asked the user to carry out a single task like, for example, *Click on cell C3*, or *Select Insert from the menu bar*. Action step 24 asked to type in a rather complex formula that was hard to memorize, which also caused a need for multiple switches.

The rank order measure indicates that perhaps the conditions are not as similar as the relative and absolute measures suggest. Friedman's rank order test revealed a statistically significant difference ($\chi^2(2) = 6.671$, $p = 0.036$) with mean ranks 2.16, 2.14, and 1.70 for the Visual-Full, Visual-Part&Full, and Textual manual, respectively. Because the mean ranks of the two visual conditions were almost identical, the two were combined into one and a subsequent Two Groups Sign test was run. This test too showed a statistically significant difference favoring the visual manuals ($z = -3.050$, $p = 0.002$). The exact comparison between the visual and textual manuals is 32 cases in which the rank of the Visual manual is higher than the Textual manual, 11 cases in which the Visual manual is lower than the Textual manual, and two cases which rank equally.

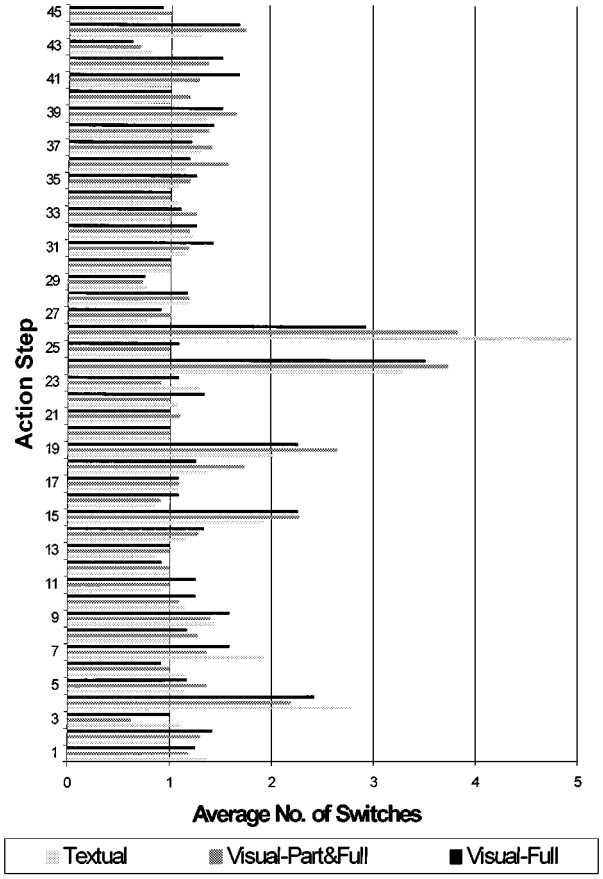
DISCUSSION

The main aim of this study was to find empirical support for the view that screen captures stimulate users to switch attention. Although the findings from this study are not entirely conclusive in this respect, based on the results of the rank-order analyses, there are indications that a manual with screen captures may stimulate users to switch more often between manual and computer screen than does a textual manual.

For the percentage of switches a ceiling effect was found. Regardless of manual type, the users switched attention in about 97% of the instructions. In other words, users in all conditions switched attention when there was an objective need to do so. This finding calls into question the added value of a visual design. Perhaps it simply indicates that the action-oriented approach of the (minimalist) tutorials in combination with the user's computer skill suffice to realize a desirable minimum of a "switching mode."

The results of this study did not lead to insight in the benefits or drawbacks of a particular visual design to support switching attention. No test pointed to a

Fig. 4. Average number of switches per action step for three manual types.



difference between design variants. An explanation for this may be that the full screen captures were very readable after all. If that were the case, both visual manuals become equal in use. There is no longer an advantage for the well readable partial screen captures of the Visual-Part&Full manual over the Visual-Full manual without those partial screens. At the same time, both designs become evenly self-contained, canceling out the advantage of the Visual-Full manual that was supposed to force the user to look up to the screen over the self-contained Visual-Part&Full manual.

A second explanation may be that users of the Visual-Part&Full manual ignored the redundant partial screen captures. If these screen captures were indeed not used, this neutralizes the difference in design between the two visual variants.

A third explanation may be that users adapt to the design of the manual. Although they equally benefit from the screen captures offered, their use of these visuals could have been different. More specifically, users of the Visual-Full manual switch regularly because they are forced to do so by the unclear screen capture in the manual in comparison with the actual screen. Users of the Visual-Part&Full manual switch equally often because the combination of partial and full screen captures makes switching easy and thus convenient for them. Although the results are the same, the realization of these results may have been different. Studies on such exact picture processing may provide further insight in the effects of screen captures for switching attention between the manual and computer screen.

ACKNOWLEDGMENT

The authors would like to thank C. Blockhuis and M. Korte for their valuable help on preparing the manuals used in this study and for carrying out the observations. They would also like to thank T. de Jong and J. Pieters for their reviews on earlier versions of this paper.

REFERENCES

[1] J. M. Carroll, "Designing minimalist training materials," *Datamation*, vol. 30, pp. 125–136, 1984.
[2] H. Van der Meij, "The ISTE approach to usability testing," *IEEE Trans. Prof. Commun.*, vol. 40, no. 3, pp. 209–223, 1997.
[3] J. Sweller and P. Chandler, "Why some material is difficult to learn," *Cognition and Instruction*, vol. 12, pp. 185–233, 1994.
[4] D. C. Gardner and G. J. Beatty, *Visuele leermethode Windows 3.1 [Windows 3.1: The Visual Learning Guide]*. Utrecht: A. W. Bruna Uitgevers B. V., 1994.
[5] M. Gellevij, H. Van der Meij, T. de Jong, and J. M. Pieters, "The effects of screen captures in manuals. A textual and two visual manuals compared," *IEEE Trans. Prof. Commun.*, vol. 42, no. 1, pp. 77–91, 1999.
[6] H. Van der Meij, "The role and design of screen images in software documentation," *J. Comput. Assist. Learning*, vol. 16, pp. 294–306, 2000.
[7] T. J. H. M. Eggen and P. F. Sanders, *Psychometrie in de praktijk [Psychometrics in Practice]*. Arnhem, The Netherlands: Cito, 1993.
[8] T. Bechger, B. Hemker, and T. Heuvelmans, *Het bepalen van beoordelaarsovereenstemming in een vooronderzoek voor de staatsexamens Nederlands als tweede taal en de Profieltoets [Determination of Interrater Reliability in an Analysis for the State Examens Dutch as a Second Language and a Profile Test]*. Arnhem, The Netherlands: CitoG, Psychometrisch Onderzoekscentrum, 2001.

Mark Gellevij works in the Department of Instructional Technology of Twente University, Enschede, The Netherlands. After graduating from this department in 1995, he worked as a technical writer in trade and industry. He is completing his doctoral thesis concerning the function of screen captures in software manuals and the cognitive processing of text–picture combinations. As a teacher, he specializes in instructional theory and design.

Hans van der Meij works in the Department of Instructional Technology of the University of Twente, Enschede, The Netherlands. His main areas of interest are information-seeking behavior and technical documentation. In the latter area, he has published 16 chapters and 33 articles. Two of his articles received a "Distinguished Article Award" from the Society for Technical Communication. In 1995, for an article coauthored by J. M. Carroll and, in 1998, for an article coauthored by Mark Gellevij, he received the "Best Transactions Paper Award for 1997" from the IEEE Professional Communication Society. His studies in technical documentation combine theory, practice, and empirical research. He currently conducts inventories on the roles and designs of screen captures, procedures, warnings, and indexes.