The effects of an active development of the mental model in the training process: experimental results in a word processing system

MICHAEL FRESE, KAREN ALBRECHT, ALEXANDRA ALTAMANN, JUTTA LANG, PATRIZIA V. PAPSTEIN, REINHARD PEYERL, JOCHEN PRÜMPER, HEIKE SCHULTE-GÖCKING, ISABEL WANKMÜLLER and RIGAS WENDEL

Department of Psychology, Ludwig-Maximilians-Universität, Leopoldstrasse 13, 8000 München 40, FR Germany

Abstract. Three different training programmes for a word processing system were experimentally compared: (1) a sequential programme, which taught low-level skills and which did not help the user actively to develop a mental model, (2) a hierarchical programme, which provided an explicit and integrated conceptual model of the system to the user and (3) a programme in which the users were asked to develop hypotheses on the functioning of the software and in which they were encouraged to use an active and exploratory approach. From an action theory point of view it was hypothesised that the third group would be superior to the first group. In an experimental study with two training sessions each of two hours and a two-hour testing session (n = 15), this was shown to be the case for several performance criteria (error time, transfer and experimenter rating). Additionally, an interindividual difference variable to measure the individual learning style was used, giving results that could be interpreted in a similar way to the experimental results.

1. Introduction

While it seems obvious that there is a need to develop software that is functional, usable and user friendly, there is relatively little emphasis on training in industry. When computers are introduced in the workplace, training issues do not usually receive a high priority (Algera et al. 1986, Bjorn-Andersen 1985, Gottschall et al. 1985).

However, there is ample evidence of the importance of training. Even 'foolproof' office systems with a desktop monitor require some form of training (Altmann 1987). They are difficult to learn even for persons who are familiar with other word processing systems (Carroll and Mack 1985). Training is necessary even if so-called tutorials or manuals are provided. Manuals are not much used (Carroll and Mack 1983, Scharer 1983) and on-line tutorials are often not helpful (Greif 1986).

Therefore, it is useful to ask the question, what type of training renders the most efficient results? It is possible to describe different training programmes with two dimensions (see table 1): the dimension of how the training process is organized and the dimension of developing the mental model (Norman 1983). The first dimension (training process) distinguishes sequential from integrated programmes. In a sequential programme, the action is partitioned into small low-level sequences which are taught and practised separately, and are then pieced together again. In word processing this would mean, for example, that a person learns the commands for deleting, for inserting, etc., and practises each one of them individually. Many computer aided tutorials follow this kind of programme, presenting a step and then asking the person to perform the appropriate action (Greif 1986).
Table 1. Development of mental model.

<table>
<thead>
<tr>
<th>Training process</th>
<th>Passive</th>
<th>Active</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequential</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integrated</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

From everyday considerations, one could assume that this type of training has the advantage of reducing anxiety in the learning process (because only one command has to be learned at a time); but it has the disadvantage that a relatively complete and integrated model of the system cannot be developed. The integrated training programme, on the other hand, furthers an integrated mental model in which high-level explanations are used and understanding of the whole system is emphasized.

The second dimension differentiates passive from active training strategies in developing a mental model of the system. In the passive programme, the trainer teaches the system’s model and the computer novice is required to learn it. In the active approach, the trainee develops a mental model of his/her own and tries out its adequacy.

Our starting point for studying the effects of different training procedures was the theory of action (Frese and Sabini 1985, Volpert 1981). It is not possible to describe this theory in detail here. The following list of five propositions are related to the theory:

1. People are active learners. Therefore, students do not typically follow prearranged programmes in detail. They usually try out certain steps and try to explore the system (Carroll and Mack 1984). Training programmes which encourage an active approach and use exploratory approaches lead to better performance than programmes which do not.

2. The action is regulated by an operative image system (Hacker 1978) or an action oriented mental model (cf. Gentner and Stevens 1983). A more integrated and coherent operative image system leads to better performance.

3. Novices start out with a rudimentary conceptualization of the system (in fact this guides their activities in the first hours of work). The more it is possible to integrate new information into these already existing conceptualizations, the better they are able to learn (Volpert 1981).

4. A person does not only learn from doing something correctly but also from making mistakes. Feedback is critical here (Annett 1969) because it provides information for developing an integrated and differentiated mental model. Since exploration and making mistakes leads to a better operative mental model, people will actually make fewer mistakes when using their skills.

5. The rules and plans that regulate actions are conscious in the beginning of the learning process. With practice they are automatized (Volpert 1981). However, when there is premature automatization, one tends to automatize inadequate behaviours and performance turns out to be worse than it should be.

Thus, our hypothesis follows from this thinking: a training programme that encourages exploration and the active development of an integrated mental model will lead to better performance. A sequential training that allows only a passive development of a mental model will be worse in comparison. This is because a sequential programme does not encourage the active development of a mental model and is less likely to produce a coherent mental model. Furthermore, it may lead to a
The mental model in the training process

premature automatization (since each individual step is practised outside the context of the other commands).

In order to test these considerations, we have developed training programmes for the word processing program Wordstar which filled three of the four cells in Table 1 (it is very difficult to develop a training programme for the fourth cell, the one combining an active development of a mental model and a sequential training, because people spontaneously develop integrated mental models). In our experiment, three training groups of novices without any prior experience with computers were compared:

1. The sequential group (passive development of mental model and sequential process). The emphasis in this group was on developing low-level skills. Thus, the students received written material which told them step by step why a certain command had to be used and no mnemonic aids were given. The chances to make errors were minimized; as long as the subjects followed the instruction sheet, they could not make any errors.

2. The hierarchical group (integrated training process and passive development of mental model). This group received a manual (23 pages long, written for this experiment) plus a hierarchical and organized diagram presenting all the commands to be learned. This material gave explanations and mnemonic aids, and presented a model of a detailed hierarchical and integrated word processing system.

3. The hypotheses group (active development plus integrated mental model). This group did not get any written material on the computer system because the subjects were supposed to develop actively their own coherent mental model. Before starting to work with the computer, they received a hard copy of the flawed text and were asked to develop hypotheses about commands to be used to correct the mistakes with the help of the computer. When using the computer these hypotheses were elaborated with the guidance of the experimenter down to the keystroke level. The subjects were encouraged to try out solutions which they had thought of. The correct command was then mentioned and the subjects could write down what they had learned.

According to our theoretical conceptualization, the first group should perform worst, the second group should be in the middle and the third group should perform best. The sequential group (first group) essentially mimics a rote learning condition; they are encouraged to learn the various commands by heart and are discouraged from developing an overall understanding (mental model) of the system. Furthermore, there may be premature automatization of the commands without their being fully understood. The second group is at least provided with a coherent conceptualization of the system but there is no effort to assure that this conceptualization fits their prior thoughts and hypotheses or is integrated with them. The subjects of the third group can develop an integrated mental model based on their own previous conceptualization. Here the subjects learn to approach the new system actively, which should also lead towards a more active approach when new tasks have to be solved. An active approach is better able to match the 'normal' adult approach to learning something new.

The differences between the groups should be more pronounced the more difficult the task is. A task is more difficult if it has to be done under speed conditions or if complicated rules have to be applied.
The advantage of a more active and exploratory approach may not only be true for the third training group but may also hold for individuals in other groups: people who show a learning style that is characterized by an active and exploratory approach. An active and exploratory learning strategy should lead to a better mental model and to a more active approach in general, regardless of the training group. In several small pilot studies, we found that there are some individuals who want to be quite sure about their knowledge before they start working on a system; therefore they read manuals more diligently. Others start immediately to interact with a computer system. The latter learning strategy should lead to a better performance since it encourages exploratory strategies and an active development of a mental model.

2. Methods
2.1. Subjects
The subjects were 15 students, five in each experimental group. The subjects were assigned at random to the three groups. They had different specialisms but they were mostly from the social sciences (business, psychology, etc.). Ten were males; five were females; and the mean age was 27 years. They were complete computer novices and quite stringent criteria were used to select them. While there were many applicants who wanted to learn the word processing system (about 45), only 17 reported having had no experience with computers. We learned during the training that two of these had had some prior experience with computers; they were therefore excluded as well.

2.2. Experimental procedure
Although the subjects were taught in groups, each subject had his/her own computer and experimenter. The goal of the training was to teach those rudimentary Wordstar commands that are needed to write a paper (moving the curser, changing disk drive, deleting, inserting, setting the margins, underlining, bold print, printing, etc.).

All subjects were trained for the same number of hours. The experiment consisted of three two-hour sessions (two training sessions and one testing session). In the first session, the subjects were trained by correcting a flawed text (they also received a correct version on hard copy). If subjects finished before the two hours were over, they were encouraged to redo the correcting. (This happened frequently in the first group but seldom in the other groups.) In the second session, on the next day, the subjects had to type a copy of a relatively complicated text, using the word processing system.

Whenever the subjects made an error, the experimenter corrected the error, intervening in the appropriate way. This meant that the experimenters had to be trained to use the appropriate rules consistently, according to the group. In the sequential group, the intervention consisted of telling the subject the correct command without further explanation. In the hierarchical information group the experimenter explained how the error had come about and in what way the error was related to the information presented in the manual and in the diagram (which displayed the hierarchy of commands). The subjects were encouraged to read the information that was necessary and it was explained to them. In the hypotheses group, the experimenters told the subjects relatively little but asked questions when the subjects made a mistake. The questions were: "Why do you think you have made an error and what do you hypothesize to be the right procedure?" The subjects were then encouraged
to try out their hypothesized solutions. Suggestions were given only when they did not get the right answers in this way.

Thus, the emphasis in the sequential group was on learning the right commands without attempting to teach a coherent mental model of the system. The emphasis in the hierarchical information group was on teaching the right model. Here the experimenters did a considerable amount of coaching. Experimenters' structuring and intervening (other than asking questions) were lowest in the hypotheses group since the emphasis there was on exploration and developing one's own mental model. No double blind procedure was used because the experimenters had to know the reasoning behind the different procedures to follow the rules correctly.

2.3. Measures

The following measurements were taken (all the performance measures were taken in the third session):

1. **Free recall.** At the beginning of the second and the third sessions, the subjects were asked to recall all the commands that they still knew and explain what they could be used for. The correct answers were counted.

2. **Corrected error score.** In the third session, the subjects were required to type a copy of a text under speed conditions. As in Roberts and Moran's (1983) benchmark tests, we ascertained how many errors they made. Each error was scored (unlike Roberts and Moran's tests, in which small errors were excluded) and the time taken to correct the error was noted. If the subject did not notice the error, the experimenter pointed it out to the subject. Time was measured with a stopwatch. When subjects did not manage to correct an error within 120 seconds, they were told to continue typing the text. Similarly to Roberts and Moran, we computed the corrected error score with the following formula: error time × 100/(overall time — error time). The correction of this score consisted of weighting the error time with overall error-free time. Fifteen minutes were allotted to this task.

3. **Inefficiency.** We also wanted to use a non-speed performance test (to provide a contrast to the corrected error score). Therefore, the subjects were asked in the third session to correct a text that contained a set of errors. The text was presented on the screen. Again, this measure was developed in the spirit of Roberts and Moran's benchmark tests. The keystrokes needed to correct each error were counted. The more keystrokes a subjects needed, the more inefficiently he/she was working. We differentiated a priori between easy tasks (inefficiency easy), tasks of moderate difficulty (inefficiency middle) and difficult tasks (inefficiency difficult). It was assumed that the differences between the groups should be higher for the difficult tasks.

4. **Transfer.** The subject had to use a command that they had not learned before in the two training sessions. The task was relatively easy (using CTRL Y to delete a line); the subjects could get the right answer by just reading the menu (however, we had not encouraged the subjects to use the menu during training). The time needed to solve this new problem was measured with a stopwatch. When subjects could not solve the problem within 10 minutes, we interrupted the task (and a 10-minute count was taken for the subject).
(5) Performance (experimenter rating). An overall performance rating was given by the experimenter on how well the subject was able to use the commands taught in the course (five-step scale).

(6) Satisfaction.
   (a) An overall subjective satisfaction was ascertained with Kunin faces (Kunin 1955) (seven-step scale).
   (b) Four questions to assess satisfaction with the training were developed for this study along the lines of the job satisfaction literature (Bruggemann et al. 1975), e.g. Would you suggest to a friend that they should participate in this training? How satisfied were you with the progress that you made during this training? These were combined to form a questionnaire. This questionnaire had an acceptable reliability of Cronbach’s alpha, \( r = 0.84 \).

(7) Learning style (learning by studying – reading the manual – versus learning by doing – active and exploratory strategy). The questionnaire developed to assess this (by Schulte-Göcking 1987 and Prümper 1987) is not directly related to computer issues (the questions refer to other tools). It asks whether one is likely to read and amass a lot of information before starting to act rather than adopt an exploratory approach to a new tool.

(8) Typewriting skill (experimenter rating). Typewriting skill may be a potentially confounding variable because the performance while speed is being measured may be affected by it. Therefore, the experimenters gave an overall rating for the typewriting skills of the subjects (using a seven-point rating scale) and this variable was used as a covariate.

3. Result and discussion

Table 2 presents the means of the dependent variables for the different experimental groups. The performance measures of how to do the tasks (numbered from 3 to 6 in the table) present a picture that is quite clear overall. In general, group 3 (the hypotheses group) fare better than the other two; group 1 (the sequential group) is worst and group 2 (hierarchical information) is in the middle (there are practically no differences between the groups with regard to 4.1., Inefficiency easy, and 4.2., (Inefficiency middle). Although the pattern of the results for the performance measures is coherent, not all of the differences are statistically significant. In the analyses, we have concentrated on a comparison of groups 1 and 3.

3.1. Recall

The overall differences between the groups with regard to recall on the second day are marginally significant (\( F(2, 12) = 3.07, p = 0.08 \)). Here, group 2 is the worst (it is also significantly different from the third group \( t = -2.38, p < 0.05 \)). Possibly, the hierarchical group gets too much information in the beginning and, therefore, suffers from information overload, leading to worse recall. However, the differences disappear in the third session (with a very slight and unimportant better result for the third group).

3.2. Error time, inefficiency and performance rating

There are large and significant differences in 3. Corrected error time, between the groups (analysis of covariance with typing skills as covariate: \( F(1, 2, 12) = 4.94, p < 0.05 \)). Group 1 is marginally significantly different from group 2 \( t = 1.87, p = 0.06 \) and significantly different from group 3 \( t = 2.29, p < 0.05 \). Apparently, the different training procedures produce an impact on how the subjects are able to deal with errors.
The mental model in the training process

Table 2. Comparison of the three groups (means differences).

<table>
<thead>
<tr>
<th>Means of the experimental groups</th>
<th>1, sequential</th>
<th>2, hierarchical</th>
<th>3, hypotheses</th>
<th>Overall SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Recall 2nd day (no. of correct commands)</td>
<td>3·60</td>
<td>2·20</td>
<td>6·00</td>
<td>2·8</td>
</tr>
<tr>
<td>2. Recall 3rd day (no. correct commands)</td>
<td>8·00</td>
<td>8·20</td>
<td>9·40</td>
<td>2·1</td>
</tr>
<tr>
<td>3. Corrected error time</td>
<td>40·40</td>
<td>21·20</td>
<td>15·80</td>
<td>17·2</td>
</tr>
<tr>
<td>4. Inefficiency overall (no. of keystrokes)</td>
<td>40·30</td>
<td>29·42</td>
<td>25·30</td>
<td>15·6</td>
</tr>
<tr>
<td>4.1. Inefficiency easy</td>
<td>7·06</td>
<td>6·90</td>
<td>5·58</td>
<td>1·9</td>
</tr>
<tr>
<td>4.2. Inefficiency middle</td>
<td>11·08</td>
<td>11·34</td>
<td>10·42</td>
<td>1·1</td>
</tr>
<tr>
<td>4.3. Inefficiency difficult</td>
<td>22·16</td>
<td>11·18</td>
<td>9·32</td>
<td>14·1</td>
</tr>
<tr>
<td>5. Performance (rating) (1–5 scale)</td>
<td>3·00</td>
<td>3·60</td>
<td>4·20</td>
<td>0·8</td>
</tr>
<tr>
<td>6. Transfer (solution time in min)</td>
<td>4·30</td>
<td>2·23</td>
<td>1·31</td>
<td>2·5</td>
</tr>
<tr>
<td>7.1. Overall satisfaction (Kunin, 1–7 scale)</td>
<td>5·80</td>
<td>6·00</td>
<td>6·00</td>
<td>0·9</td>
</tr>
<tr>
<td>7.2. Satisfaction (questionnaire, 1–5 scale)</td>
<td>3·80</td>
<td>4·30</td>
<td>4·20</td>
<td>0·7</td>
</tr>
</tbody>
</table>

While there are similar differences between the groups for the dependent variable 4, Inefficiency overall, the difference between groups 1 and 3 is only marginally significant ($t = 1·93, p = 0·07$). In line with our hypothesis, the difference is almost entirely due to inefficiency in difficult tasks (4.3., Inefficiency difficult: $t = 1·48, p = 0·09$). Thus, differences between the groups tend to show up in difficult tasks, either in terms of complexity (as in 4.3., Inefficiency difficult) or in terms of speed (as in 3., Corrected error time).

The differences between groups 1 and 3 for variable 5., Performance (rating), are significant ($t = 2·45, p < 0·05$). While one could argue that this is due to experimenter bias (since they knew the hypotheses of this study), this difference is in line with the results of the other performance measures and, therefore, lends credibility to the overall effects of the training procedures.

3.3. Transfer

There is a significant difference between groups 1 and 3 for 6., Transfer (how long the subjects needed to solve the transfer task) ($t = 1·76, p < 0·05$). Again, the hypotheses group is superior in performance.

3.4. Satisfaction

The two measures of satisfaction, 7.1. and 7.2., do not show any important differences between the groups. This reproduces other results showing that subjective satisfaction and attitude measures are not significantly correlated with performance (Barnard et al. 1981). This result is plausible because the subjects lack the knowledge of other training methods and are, therefore, not able to compare them. Most trainees were quite satisfied with their specific training and their trainers. This also showed up in
some qualitative data; we asked the subjects to describe a method by which they would teach a friend the word processing programme. The suggestions most often mimicked the respective method under which the subjects were trained in the experiment.

3.5. Interindividual differences

We hypothesized that interindividual differences might have a similar influence to the different training procedures. When individuals use less active and exploratory strategies in general, they develop a less coherent and integrated mental model, regardless of how they are trained. (Actually, we also hypothesized an interaction of the training procedure with learning style but this could not be tested in this experiment owing to the small number of subjects.)

Table 3 presents the correlations of learning style (high score = inefficient strategy of learning by studying) with the dependent variables of this study. The overall picture is consistent with the data from the experiment. Again, there is the exception of the recall measures (recall on the third day is non-significantly and positively correlated). There are relationships with the performance variables – albeit many of them not significant or only marginally significant. Apparently, when learning to use a system such as Wordstar, it helps to have an explorative and active learning style.

Table 3. Correlations of learning by studying with the dependent variables.

<table>
<thead>
<tr>
<th>Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Recall 2nd day</td>
</tr>
<tr>
<td>2. Recall 3rd day</td>
</tr>
<tr>
<td>3. Corrected error time</td>
</tr>
<tr>
<td>4. Inefficiency overall</td>
</tr>
<tr>
<td>4.1. Inefficiency easy</td>
</tr>
<tr>
<td>4.2. Inefficiency middle</td>
</tr>
<tr>
<td>4.3. Inefficiency difficult</td>
</tr>
<tr>
<td>5. Performance (rating)</td>
</tr>
<tr>
<td>6. Transfer time</td>
</tr>
<tr>
<td>7.1. Overall satisfaction</td>
</tr>
<tr>
<td>7.2. Satisfaction</td>
</tr>
</tbody>
</table>

* p < 0.10; ** p < 0.01; † p < 0.05.

4. Overall discussion

There are some limitations to this experiment. It was based on a rather small number of subjects; we did not use very conservative statistical tests; and we could not control the sex of trainee–trainer pairs (there is some evidence that subjects with a trainer of the opposite sex showed better performance results than those with a same sex trainer, but because of the small number of subjects we could not pursue this lead). However, our general interpretation can be justified by the data, since all the performance results point in the same direction. The overall picture is persuasive, although not all the specific results reached traditional significance criteria. While we cannot say much about the hierarchical information group, the results suggest that the sequential training programme with its emphasis on rote learning of commands is inferior to the programme used for the hypotheses group. This speaks for a training programme that demands an active development of an integrated mental model.
Interestingly, the active approach proves to be superior not only as a result of the experimental procedure but also as a result of a personal learning style. Novices who do not accumulate a lot of knowledge by looking at the manuals but are active and exploring when learning a new computerized system achieve better results. We do not suggest that such a learning style (learning by doing) will always be superior. Its superiority may only appear when the training material is relatively easy to master. (Perhaps it is only then that an active approach leads to better performance.) If the material to be mastered is complicated, learning by studying may be advantageous (see Prümper 1987).

In the experimental results, recall differs from the other performance variables. While there were differences in recall on the second day, there were no longer any differences in recall on the third day. This shows that the performance effects are not due to memory differences between the groups—the performance differences appeared on the third day. This suggests that the performance differences are due to differences in how the mental models are organized. In other words, recall may tap mainly declarative knowledge, and the superiority of the third group may be due to a better development of procedural knowledge (Anderson 1983) or to use of the operative image system. The use of hypotheses in the beginning of the training emphasizes the development of some explicit cognitive preconception of the system before one actually starts to interact with it. The newly learned material can be integrated into these preconceptions. The preconceptions can then be elaborated into a full mental model. Although we do not have an adequate independent measurement of mental models, the performance results suggest that the mental models of the subjects in the hypotheses group are more flexible, more integrated, better anchored in the subjects’ preconceptions and encompass more knowledge of errors.

If the results of this study can be cross-validated in other studies, it would mean for practical purposes that emphasis should be placed on the self-developed mental models in the training process. Brown and Newman (1985) argue in general for exploratory and active strategies in learning that should be supported by the appropriate software. Our results are in agreement with their claim and support their line of reasoning. There is little reason for continuing with a non-thinking, rote-learning paradigm in what is one of the most cognitive, thought-involving, integrative and active activities of mankind—learning.

Acknowledgments
We are grateful to the Leibniz Rechenzentrum for their support of this research project. Thanks are due to F. Glover for commenting on an earlier draft of this article and to U. Jäger who helped in the data collection. Requests for reprints should be sent to Michael Frese, Department of Psychology, Leopoldstrasse 13, 8000 München 40, Federal Republic of Germany.

References
The mental model in the training process

GOTTSCHAL, K., MICKLER, O., and NEUBERT, J., 1985, Computerunterstützte Verwaltung (Frankfurt: Campus).
SCHARER, L. L., 1983, User training: less is more. Datamation, 29, 175–182.

Quelle: