Effect of error information in tutorial documentation

Ard W. Lazonder and Hans van der Meij

In learning to use software, people spend at least 30% of their time on dealing with errors. It could therefore be desirable to exploit users' errors rather than to avoid them. That is, to include error information in a manual to support users in dealing with errors. An experiment was performed to examine the functionality of such error information in a manual for a word processor. Two manuals were compared, one with error information and one from which this information was removed. Forty-two subjects were randomly assigned to one of the two conditions. Subjects who used the manual with error information were expected to become more proficient at using the word processor (i.e. to show better constructive and corrective skills) and to develop more self-confidence. The results were equivocal. On some aspects of skill the error information in the manual led to better performance (i.e. correcting syntactic errors). On others it had an adverse effect (i.e. detection of semantic errors and overall error-correction time). Explanations are advanced for these findings and topics for further research are identified.

keywords: human-computer interaction, human error, tutorial documentation, error information, corrective skills development

One of the most striking features of first-time computer users is that they are active learners. They want to do things in order to reach personal goals, rather than read through endless pages of what they consider to be 'just information' (Carroll, 1990b; Redish, 1988; Scharer, 1983; Wendel and Frese, 1987). Unfortunately, the instructional strategy of most computer manuals does not suit this spontaneous learning strategy. Most manuals require users to proceed step-by-step through endless series of drill and practice exercises, giving them (too) little freedom for active (i.e., self-initiated) learning.

The main reason for this is that trainees get themselves in trouble when they explore (Njoo and De Jong, 1991). Although these explorations can be advantageous in learning computer-related tasks (e.g. Kamouri et al., 1986; Van

Department Instructional Technology, University of Twente, P.O. Box 217, 7500 AE Enschede, the Netherlands. Tel: X-31-53-893596. E-mail: lazonder@edte.utwente.nl

Paper received: January 1993; revised: August 1993

Jooringen, 1993), this view is not generally accepted. Consequently, detailed step-by-step instruction is supposed to prevent users from making mistakes. This assumption is unrealistic, however. Research (e.g., Carroll and Mack, 1984; Mack et al., 1987; Redish, 1988) has consistently shown that new users have problems following the descriptions that manuals provide. Frequently, they consider these directions paradoxical or irrelevant to their goals. As a consequence, novice users tend to explore the system on their own. As many as 65% of the users may skip information they consider irrelevant and use the manual only when they need help (Penrose and Seiford, 1988).

When novice users ‘jump the gun’, problems can and will arise. Just as in step-by-step instruction, mistakes occur during exploratory behaviour. That is, the user may be blocked from any further exploratory actions. Research has consistently indicated that new users spend 30-50% of their time on detection and correction of errors (Bailey, 1983; Card et al., 1983; Graesser and Murray, 1990).

Errors and learning

In conventional manuals, errors are seen as ‘blocks’ to learning that can be avoided by detailed step-by-step instruction. This view is in line with classical theories (e.g., behaviourism) that aim for a minimum number of errors. These learning theories advocate error minimization because:

- errors hinder the control over learning;
- errors cause frustration, which, in the end, may cause a learner to stop learning;
- errorless learning is richer in association, because it prompts and explicitly relates new knowledge to existing knowledge (Glaser, 1965; see also Glaser and Bassok, 1989).

A contrasting, and more fruitful approach in view of the above, is to perceive errors as a wonderful opportunity for learning (e.g., Brown et al., 1982; Carroll, 1990a; Singer, 1978). There are two reasons for this. First, the nature of the learning experiences should reflect the intended training outcomes as much as possible. Learning to master a system means developing constructive and corrective skills. Users must learn not only how to do things, but also how to undo things that go wrong. That is, they should also learn how to deal with errors. Training should therefore focus on the development of both these procedural skills (Wendel and Frese, 1987). Second, errors signal misconceptions in the users’ conceptual model (Brown et al., 1982; Pickthorne, 1983; Stevens et al., 1982). Errors may thus help users to reveal and remove these misconceptions, and, consequently, develop a better conceptual model.

Errors will only have a positive effect if they are controlled in the learning process. That is, when the instruction supports the user’s corrective skills development. This paper investigates how such error control can be brought about. It first outlines the stages involved in dealing with an error. From this model, demands for effective error control are identified. The second part of this
paper reports an experiment that tests whether a manual that meets these demands assists first-time users in developing word processing skills.

**General model of error-recovery**

The main goal of users who have made an error is to return to a normal, or at least acceptable system state (Johannsen, 1988). In achieving this goal, users tend to go through three stages: detection, diagnosis, and correction (Brown, 1983; Curry, 1981; Jelsma and Bijlstra, 1990; Wærn, 1991). The main assumption behind these stages is that all user-activity is goal-directed (cf. Ashcraft, 1989; Card et al., 1983; Norman, 1986; Stillings et al., 1987). A detailed outline of these stages is presented here.

**Detection**

Error detection is the first step in recovery. It is a precondition for the other stages: without detection, users will not attempt to correct mistakes.

An error is detected when a user considers an outcome to contrast with his or her original goal. More specifically, there are two ways in which error detection may be triggered (Allwood, 1984). First, triggering comes in response to some external cue. For example, a user perceives a discrepancy between an outcome and some definitive yardstick of correctness (Guthrie et al., 1991; Lewis, 1981). Second, detection can be prompted internally. That is, it can be initiated by the user on his or her own accord (Lewis, 1981). The user may, for example, feel insecure with the selected method, the command(s), or its execution.

Triggering is not a sufficient condition for detection. The user may, for example, abandon the pursuit of an error that is not important and does not interfere with task execution. Moreover, triggering does not always occur at the right moment. Misconceptions about the expected outcome, or the appropriateness of a solution method may lead to undetected errors or to a delay in the detection of an error. On the other hand, triggering can also occur if no error has been made. In that case, correct performance is judged as erroneous.

So, in addition to triggering occurring, the user has to spot where the error is to actually detect it. Locating an error occurs by evaluating, or reviewing the current system state and the actions that were performed.

**Diagnosis**

After detection, the nature of the error is still only vaguely known. The user merely knows that something has gone wrong. In diagnosis, the two main activities are finding out the exact nature of the error and its possible cause.

First, the error must be identified to understand its exact nature. That is, the system's error-state must be compared with the user's original goal. By comparison the discrepancy between the observed and the desired output becomes clear. Second, the user is likely to consider what may have caused the error (McCoy et al., 1986). In the case of a more fundamental mistake, the user will wonder about the method that was applied.

Whereas the diagnosis of the nature of an error is conditional to correction, the diagnosis of its cause is not always needed for correction (Rasmussen, 1986). It does help users develop a better conceptual model, however.
Correction
Correction contains four different kinds of user activity. First, the user must select a (repair) goal. Having identified the difference between where the user is now and where he or she wants to be, the goal is obvious. The gap between the actual and desired output (i.e., the user's original goal) must be bridged. This is often done by subgoal decomposition (e.g., Anderson, 1985; Frederiksen, 1984; Newell and Simon, 1972). For example, the user may divide the overall goal 'correct a typo' into three subgoals: move the cursor, delete the incorrect text, and type the correct text.

Second, the user must plan the method, for there may be more than one method to achieve the repair goal. To select the most appropriate method, the users decide which selection rules apply (Card et al., 1983). Each of these rules has the form of an 'if-then' statement. In the above example, one of the selection rules for cursor movements might be: "if the document contains 1 page, then use the arrow-keys to move the cursor; if the document contains 2 pages or more, then use the search-command".

Next, the method is translated into a physical action-sequence. The user selects the commands that will be used and determines in which order they will be executed. The last action in the model is the execution of the commands. Execution is the first physical action in this model.

Errors may be given different statuses. Some errors will be easier to detect and/or correct than others. For that reason, errors are classified into one of the following categories: semantic, syntactic, and slip (cf. Douglas and Moran, 1984; Lewis and Norman, 1986). A semantic error occurs when an inadequate command is chosen to achieve a given goal. For example, the user may select 'Base Font' to try to set a word in italics. When a correct command is carried out improperly, it is called a syntactic error (e.g., changing the line spacing into 1½ instead of 1.5). Slips are small mistakes at the keystroke level (e.g., typing errors)*. In general, there is no research on how to deal with these three types of error in computer documentation. Therefore, this study will only explore any differential effects.

Toward effective error control
To bring about effective error control, a manual should support users in dealing with errors. Such control is possible by including error information in the manual. In keeping with the staged error-recovery model, good error information should consist of:

- a characterization of the system-state to detect and identify the error,
- conceptual information about the likely cause of the error,
- action statements for correcting the error (Lang et al., 1981; Mizokawa and Levin, 1988; Roush, 1992).

*One of the reviewers also suggested considering task-level errors (e.g., grammatical errors, punctuation errors). We agree with this notion. We also feel that these mistakes cannot be supported by the error information in the manual. They are therefore not further discussed.
Retrieving a document

Before you can retrieve a document from disk, you must always clear the screen first.

1. Go to the menubar and choose the command EXIT
2. Press the ENTER key
3. Answer both questions by typing an X

You have cleared the screen.

If there is still text on the screen, you may have pressed the wrong key. Press the F7 key and type an X twice to clear the screen as yet.

4. Go to the menubar and choose the command RETRIEVE
5. Press the ENTER key

If the text Document to be retrieved: does not appear, you have selected the wrong command. Press the F1 key to rectify your choice.

6. Type MANUAL.TXT
7. Press the ENTER key

The document MANUAL.TXT appears on the screen

If the screen remains empty, you have probably made a typing mistake. Retype the name of the document and press the ENTER key.

Figure 1. Error information in a manual. Left-hand column shows an example page of the manual with error information (MM+) that was used in the experiment. Right-hand column shows the corresponding principles for effective control.

A typical example of error information might thus read:

“If the code [Hrt] appears, you pressed the RETURN key instead of the F2 key. Remove the [Hrt] code by pressing the BACKSPACE key. Press the F2 key to start the search as yet.”

Special attention should also be given to the appropriate timing of the error information in order to reduce the number of delayed detections. Error information should be presented frequently, often directly after the commands, rather than in separate ‘trouble-shooting’ sections (Bailey, 1983; Carroll, 1990b; Horton, 1990; Lewis and Norman, 1986). As a rule of thumb, it is to be presented when errors have severe consequences for further task execution, or when commands are error-prone. An example of how error information should be incorporated in a manual is shown in Figure 1.

To examine the efficacy of error information, an experiment was conducted using a manual with error information (MM+) and a control manual containing no error information (MM–). It was expected that subjects who used a manual with error information would develop better procedural skills than subjects who used a manual without error information. More specifically, the MM+ subjects were expected to perform better on test items measuring constructive and corrective skills. Constructive skills are needed to achieve the user’s
original goals, whereas corrective skills are necessary to recover from errors (i.e., meet the repair goal). Error information provides users with a safety net (Brown, 1983; Carroll and Mack, 1984; Cuff, 1980). It assures them that, no matter how odd the system's response may seem, nothing is wrong as long as the described error-state does not occur and that possible errors can be corrected at all times. MM+ subjects were therefore expected to become more confident as well.

Method

Subjects
The experiment was part of an introductory computer course for first-year students in Instructional Technology. Forty-two students took part in the experiment, receiving course credits for participation. There were 10 males and 32 females with a mean age of 19.0 (SD=1.34). Subjects were randomly assigned to one of the two experimental conditions. There were 21 subjects in the MM+ group and 21 in the MM- group. All subjects had some experience with computers (games and/or applications), but very little or no experience with the software used in the experiment. Preliminary checks on the random allocation to conditions revealed no significant difference between the two groups with regard to age, sex, educational background and initial self-confidence. The mean prior experience with computers was equal for both groups as well.

Materials

Experimental setting and word processor
All sessions took place in a computer class provided with a network of 19 Sirex 386-SX personal computers. The goal of the course was to teach elementary word-processing skills with the menu-driven version of WordPerfect 5.1. WordPerfect was downloaded from the network, thereby assuring an identical set-up of the word processor for all subjects.

A registration program was installed on each computer. It stored the subjects' actions in a logfile. Every time a key was struck, time and keypress were recorded.

Instructional materials
Subjects from both groups received a manual (MM+ or MM-) and a diskette containing all documents to be used in practice. Both manuals were so called minimal manuals, designed especially for the experiment. The two manuals varied only with regard to error information. In the MM+, all error-information was designed according to the criteria for effective error control (see Figure 1). The MM- contained no error information at all. It introduced the two main function keys for error-recovery in the first chapter, however. A more detailed description of the MM+ can be found in Carroll (1990b), Lazonder and Van der Meij (1992, in press) and Van der Meij (1992).
Questionnaires and tests
A background questionnaire was used to gather some personal data such as age, sex, educational background, and computer experience.

Subjects' confidence was assessed by three questionnaires. The first questionnaire determined subjects' initial confidence; the second and third assessed their confidence after practice and after the tests, respectively. Each questionnaire contained 20 behavioural descriptions, nine of which were fillers. The subjects judged each description (e.g., "Working with computers scares me") on a 5-point agree-disagree scale. Pilot studies revealed satisfactory reliability scores for the questionnaires (Cronbach's alpha ≥0.90).

Three tests were administered to assess learning outcomes. One test measured the subjects' constructive skill. It contained five retention tasks (i.e., elementary word-processing skills rehearsed during practice, such as removing text or changing the line spacing) and five transfer tasks (i.e., tasks not covered by the manuals, such as changing the position of the page number or adjusting the margins).

Two tests assessed the subjects' capacities for error-recovery: a knowledge and a skill test. The corrective knowledge test was a paper and pencil test. It contained three semantic errors, five syntactic errors and one slip. Each item presented a goal and a screen-dump, displaying the result of a set of actions to achieve that goal. For each item, the subjects had to mark all errors. For each error detected, its diagnosis and correction had to be specified as well. The corrective skill test was performed on the computer. Subjects had to detect and correct six errors (4 semantic, 2 syntactic) in a task document. Items of both tests were further classified into retention (i.e., included in the error information) and transfer (i.e., not covered by the error information). Since manual type had no effect on these retention and transfer scores, these measures will not be reported.

Procedure
The experiment was conducted in four groups of 7-13 subjects. In each group, half of the subjects were given a MM+ manual; the other half received a MM- manual. Separate seatings prevented interactions between MM+ and MM- subjects in a session. Within two weeks, all subjects attended two sessions of four hours each. In all, there were up to four and a half hours (maximally) for practice. The remaining time was used to complete the tests. The maximum time between sessions was three days. All procedures were identical for the various groups and the same two experimenters conducted all sessions.

At the outset of the first session, subjects filled in the background questionnaire and the initial confidence questionnaire. Next, they received instructions. The subjects were told to work individually and to consult the experimenter only when a system error occurred or when they were stuck for more than 15 minutes. They were told to work in their own way and at their own pace. The subjects were asked not to work with WordPerfect between sessions. Checks indicated that they complied with this request. After the instruction, the subjects started practice.

The second session started with another hour of hands-on experience,
enabling all subjects to complete their practise. Directly after practise, the subjects filled in the second confidence questionnaire. After a short break, they were given the constructive skill test and corrective skill test using a counterbalanced administration to control for order effects. After these tests, the subjects completed the corrective knowledge test. The subjects worked individually on all tests. They were not allowed to consult their manual or the experimenter. Enough time was given for all subjects to complete each test. After the tests, subjects filled in the final confidence questionnaire.

Coding and scoring of the dependent variables
The dependent variables were constructive skill, corrective skill and confidence. Constructive skill was defined by three measures: test time, success rate and number of errors. Test time was defined as the time required to complete the constructive skill test. A difference was made between retention and transfer. Success was indicated by the number of successfully completed items on the constructive skill test. This was assessed by examining the task documents stored on diskette and the log-files produced by each subject. The number of errors was registered for each item of the constructive skill test.

The three measures of error-recovery skill: detection, diagnosis, and correction were scored as follows. Detection was scored on a 2-point right–wrong scale. The inter-rater reliability for detection was high (Cohen's Kappa = 0.94). Diagnosis was scored on the following 4-point ordinal scale:

- both cause and effect are incorrect,
- wrong cause, right effect,
- right cause, wrong effect,
- both cause and effect are correct.

In a similar fashion, the correction method was scored as one that:

- obviously does not try to correct the error,
- attempts to correct the error, but is both semantically and syntactically incorrect or incomplete,
- is semantically correct, but contains one or more syntactic errors,
- is both semantically and syntactically correct.

Inter-rater reliability scores for diagnosis and correction were 0.77 and 0.93, respectively. For each subject, the time to complete the corrective skill test was recorded as well.

The three confidence questionnaires used a 5-point Likert-type scale. Scores on all items were added for each subject with high scores representing high confidence. Confidence changes were examined within subjects.

Results

Constructive skill
The data for constructive skill was analysed using univariate analysis of
Table 1. Mean test time scores (in minutes)

<table>
<thead>
<tr>
<th>Condition</th>
<th>MM+</th>
<th>MM-</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Retention</td>
<td>10.80</td>
<td>8.51</td>
</tr>
<tr>
<td>Transfer</td>
<td>32.31</td>
<td>21.17</td>
</tr>
<tr>
<td>Total*</td>
<td>50.26</td>
<td>21.83</td>
</tr>
</tbody>
</table>

Note: There were 5 retention and 5 transfer items.

As the time between tasks could not be taken into account for these measures, the overall test time is higher than the time for the distinct item types.

variance (ANOVA) using manual type as an independent variable. All outcomes were corrected for subjects' prior experience with computers by inserting this measure into the analyses as a covariate. Given the relative small sample size, effects of manual type on users with identical computer experience were not computed.

Time
Table 1 shows the mean time (in minutes) subjects required to complete the constructive skill test. Manual type produced no significant effect on this measure ($F(1,38)=0.41$). MM+ required as much time for completing the constructive skill test as MM- users. Retention and transfer items were also analysed separately. There was no effect of manual type on the time to complete retention items ($F(1,38)=0.64$) and transfer items ($F(1,38)=0.39$).

Quality of performance
Table 2 reports performance success. There was no significant effect of manual type on performance success ($F(1,39)=0.09$). Overall, MM+ users produced as many correction solutions as their MM- counterparts. There was also no difference in performance success on retention items ($F(1,39)=1.00$) and transfer items ($F(1,39)=0.53$).

Time and the number of successfully completed test items were combined into a measure of performance efficiency. The mean efficiency scores are presented in Table 2. As can be seen from this Table, the mean scores show no significant difference between the two groups ($F(1,36)=0.47$). Clearly, users from both experimental groups performed equally efficiently. Efficiency scores on retention and transfer items were slightly higher for MM+ users. However, none of these differences were significant at the 0.05 level ($F(1,37)=0.05$, $F(1,37)=0.06$).

Errors
The error-rates of both groups were examined by comparing the mean number of errors to the number of successfully completed items. Again, a difference
Table 2. Mean performance success and performance efficiency scores

<table>
<thead>
<tr>
<th>Condition</th>
<th>MM+ M</th>
<th>SD</th>
<th>MM− M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retention</td>
<td>3.71</td>
<td>1.19</td>
<td>4.00</td>
<td>0.76</td>
</tr>
<tr>
<td>Transfer</td>
<td>1.86</td>
<td>1.35</td>
<td>1.62</td>
<td>1.02</td>
</tr>
<tr>
<td>Total</td>
<td>5.52</td>
<td>2.06</td>
<td>5.67</td>
<td>1.39</td>
</tr>
</tbody>
</table>

*Performance success*
- Retention
- Transfer
- Total

*Performance efficiency†*
- Retention
- Transfer
- Total

Note: There were 5 retention and 5 transfer items.
*Number of items successfully completed.
†Number of items successfully completed per time (min.) × 100.

between retention and transfer items was made. The mean error-rates are shown in Table 3.

Manual type had no significant effect on the total number of errors \((F(1,38)=0.01\), indicating that, overall, subjects in the MM+ group made as many errors as subjects in the MM− group. As the mean error-rates show, MM+ users committed as many errors as MM− users on retention and transfer items. Again, manual type had no effect on the number of errors on retention items \((F(1,38)=0.65\) and transfer items \((F(1,38)=0.01\).

Corrective skill

There were three measures to assess error-recovery: detection, diagnosis and correction. Again, the outcomes were corrected for subjects’ prior experience with computers.

Table 3. Mean number of errors on correctly solved test items

<table>
<thead>
<tr>
<th>Condition</th>
<th>MM+ M</th>
<th>SD</th>
<th>MM− M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retention</td>
<td>0.48</td>
<td>0.51</td>
<td>0.62</td>
<td>0.57</td>
</tr>
<tr>
<td>Transfer</td>
<td>2.71</td>
<td>2.86</td>
<td>2.62</td>
<td>3.06</td>
</tr>
<tr>
<td>Total</td>
<td>3.19</td>
<td>2.81</td>
<td>3.24</td>
<td>3.01</td>
</tr>
</tbody>
</table>

Note: Scores are the number of errors by the number of successfully completed test items.
Table 4. Mean number of detected errors

<table>
<thead>
<tr>
<th>Error type</th>
<th>Condition</th>
<th>MM+</th>
<th>SD</th>
<th>MM−</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Corrective Knowledge Test</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semantic</td>
<td>M</td>
<td>1.86</td>
<td>0.79</td>
<td>1.48</td>
<td>1.03</td>
</tr>
<tr>
<td>Syntactic</td>
<td>M</td>
<td>2.48</td>
<td>0.98</td>
<td>2.14</td>
<td>0.91</td>
</tr>
<tr>
<td>Slip</td>
<td>M</td>
<td>0.52</td>
<td>0.51</td>
<td>0.43</td>
<td>0.51</td>
</tr>
<tr>
<td>Total</td>
<td>M</td>
<td>4.86</td>
<td>1.56</td>
<td>4.05</td>
<td>1.66</td>
</tr>
</tbody>
</table>

| **Corrective Skill Test**   |           |       |      |       |      |
| Semantic                  | M         | 3.67  | 0.66 | 4.00* | 0.00 |
| Syntactic                 | M         | 1.67  | 0.48 | 1.57  | 0.51 |
| Total                     | M         | 5.24  | 1.18 | 5.52  | 0.60 |

Note: Maximum CKT score=9, maximum CST score=6
*p<0.05

Detection
The number of detected errors were recorded by the corrective knowledge test and the corrective skill test. The mean number of detected errors are presented in Table 4.

As the mean detection scores show, the MM+ users detected more errors than MM− users. However, this difference is not statistically significant: a multivariate analysis of variance (MANOVA) on manual type by the number of detected semantic errors, syntactic errors and slips showed no multivariate effect, $F(3,37)=0.92$.

On the corrective skill test, a $t$-test on the total number of detected errors by manual type produced no significant effect, $t(40)=-0.99$. Again, the MM− group detected as many errors as the MM+ group. Manual type did affect the detection of semantic errors ($t(40)=-2.32, p<0.05$). But, contrary to expectations, the MM− users detected more errors than the MM+ users. No effect of manual type on the number of detected syntactic errors was found ($F(1,39)=0.06$).

Diagnosis
To examine the quality of the diagnoses on the corrective knowledge test, Mann–Whitney U-tests were conducted, using manual type as independent measure and the (ordinal) diagnostic scores as dependent measures. Table 5 shows the mean ranks.

Overall, there was no difference in diagnosis-scores between the two groups ($U(42)=172.5$). Apparently, the quality of the diagnoses of the MM− users was equal to that of the MM+ group. As the mean ranks in Table 5 indicate, the two groups hardly differed with respect to their diagnoses on the distinct error-
Table 5. Mean rank scores of the quality of the diagnoses

<table>
<thead>
<tr>
<th>Condition</th>
<th>MM+</th>
<th>MM-</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Error-type</strong></td>
<td><strong>Mean rank</strong></td>
<td><strong>Mean rank</strong></td>
</tr>
<tr>
<td>Semantic</td>
<td>20.48</td>
<td>22.52</td>
</tr>
<tr>
<td>Syntactic</td>
<td>21.69</td>
<td>21.31</td>
</tr>
<tr>
<td>Slip</td>
<td>22.50</td>
<td>20.50</td>
</tr>
<tr>
<td>Total</td>
<td>19.21</td>
<td>23.79</td>
</tr>
</tbody>
</table>

*Higher rank means higher quality.

Note: Diagnoses were registered on the Corrective Knowledge Test only. \( n = 21 \) for both conditions.

Table 6. Mean rank scores of the quality of the correction

<table>
<thead>
<tr>
<th>Condition</th>
<th>MM+</th>
<th>MM-</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Error-type</strong></td>
<td><strong>Mean rank</strong></td>
<td><strong>Mean rank</strong></td>
</tr>
<tr>
<td><strong>Corrective Knowledge Test</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semantic</td>
<td>20.17</td>
<td>22.83</td>
</tr>
<tr>
<td>Syntactic</td>
<td>25.00*</td>
<td>18.00</td>
</tr>
<tr>
<td>Slip</td>
<td>22.10</td>
<td>20.90</td>
</tr>
<tr>
<td>Total</td>
<td>21.79</td>
<td>21.21</td>
</tr>
<tr>
<td><strong>Corrective Skill Test</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semantic</td>
<td>20.98</td>
<td>22.02</td>
</tr>
<tr>
<td>Syntactic</td>
<td>21.00</td>
<td>22.00</td>
</tr>
<tr>
<td>Total</td>
<td>21.33</td>
<td>21.67</td>
</tr>
</tbody>
</table>

*\( p < 0.05 \)

Note: \( n = 21 \) for both conditions.

Interaction with Computers vol 6 no 1 (1994)
Table 7. Mean time (in minutes) and recovery efficiency scores on Corrective Skill Test

<table>
<thead>
<tr>
<th>Condition</th>
<th>MM+</th>
<th>MM−</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scores</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Time</td>
<td>23.30*</td>
<td>12.40</td>
</tr>
<tr>
<td>Efficiency</td>
<td>18.20*</td>
<td>9.80</td>
</tr>
</tbody>
</table>

*Mean correction scores per time × 100.

that the MM+ users were better at correcting syntactic errors than their MM− counterparts. No effect was found on correction of semantic errors (U(42)=192.5) and slips (U(42)=208.0).

On the corrective skill test, again no difference on the total correction score was found (U(42)=217.0). Apparently, MM+ users were as good at correcting errors as MM− users. The mean rank scores indicate that there was no significant difference between correction scores on semantic errors (U(42)=209.5), and syntactic errors (U(42)=210.0).

Table 7 presents the mean time (in minutes) that subject required to complete the corrective skill test. Overall, MM− users completed this test more than 7 minutes faster than MM+ users. This difference was statistically significant, t(40)=2.37, p<0.05. Time and the correction score were combined into a measure of correction-efficiency (see Table 7). Although the MM+ users were expected to be more efficient with respect to this measure, the opposite turned out to be true. Manual type had a significant effect on correction-efficiency, U(42)=127.0, p<0.05, indicating that the MM− group corrected errors more efficiently than the MM+ group.

Confidence

Error-information provides users with a safety net. Therefore, MM+ users were expected to gain more self-confidence than MM− users. Within-subject confidence changes were analysed by Wilcoxon Matched-Pairs Signed-Rank tests. The mean differences in self-confidence are presented in Table 8.

In the MM+ group, confidence scores remained relatively constant. There was a small increase in confidence after practice, and a small decrease after the tests. None of these differences were statistically significant. Confidence changes were similar for MM− users. However, the difference between confidence scores after practice and after the tests was significant for this group (Z(19)=-2.63, p<0.01). The MM− users’ confidence after the tests was lower than after practice.

Lazonder and van der Meij

35
Table 8. Within-subject differences in self-confidence

<table>
<thead>
<tr>
<th>Condition</th>
<th>MM+</th>
<th>MM-</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>B¹−A¹†</td>
<td>0.05</td>
<td>0.56</td>
</tr>
<tr>
<td>B−A₂‡</td>
<td>−0.09</td>
<td>0.69</td>
</tr>
<tr>
<td>A₁−A₂</td>
<td>−0.10</td>
<td>0.52</td>
</tr>
</tbody>
</table>

¹Confidence before training.
‡Confidence after training.
†Confidence after the tests.
*p<0.01

Discussion

This study examined the effect of error information on users' procedural skills and levels of self-confidence. Subjects who used a manual with error information were expected to develop better constructive and corrective skills and to gain a higher level of self-confidence than subjects who used a manual without error information. In general, there is no effect of error information on these measures. However, some results reveal new and interesting insights into how error information might affect user behaviour.

The first hypothesis, which stated that MM+ users would develop better constructive skills, was not supported by the results. Subjects from both conditions performed equally well on the constructive skill test. There was no difference between the two groups regarding the time to complete the test items, the number of items successfully completed or the number of errors.

Why didn't the MM+ have a facilitative effect on subjects' constructive skill? First, subjects' errors on the constructive skill test were not the kind of errors addressed by the error information in the manual. Most error information in the MM+ deals with syntactic errors. The MM's short chapters and action-oriented headings explicitly denote when commands have to be used. Information to recover from semantic errors (i.e., the choice of an incorrect command) is therefore hardly ever presented. Post-hoc analysis of the constructive skill test indicated that no fewer than 84% of the errors subjects made were semantic errors. Only 15% of the errors were syntactic; 1% were slips. Since subjects' errors were for the most part not overcome by the error information, MM+ users were not better trained to detect and correct most of their own errors. Consequently, MM+ users were not faster in completing the constructive skill test and produced as many correct solutions on test items as MM− users.

Second, the functionality of error information is affected by its actual use. During practice, subjects can use error information to correct an error or to explore the effect of a proposed correction method (see Van der Meij, 1992). This study provides no information as to whether the subjects have consulted the
error information. Subjects may not have made a given error, or they may not have explored the correction method.

The MM+ was further expected to be superior to the MM− for corrective skill. This hypothesis too was not supported by the results. On the corrective knowledge test, the MM− users detected as many errors as MM+ users. On the corrective skill test, the MM− group was faster and detected more semantic errors. Moreover, the two groups were equally proficient at diagnosing the cause of an error. With regard to correction, again no pronounced difference between the groups occurred on the two tests. The MM+ users were better at correcting syntactic errors on the corrective knowledge test, whereas the MM− users achieved higher correction efficiency scores on the corrective skill test.

The reason why MM+ users were better at correcting syntactic errors can be accounted for by the error information. As error information mainly addressed syntactic errors, MM+ users were better trained in correcting these errors than MM− users. The fact that the scores on the corrective skill test do not support this explanation can be ascribed to the correction methods used on this test. Nearly all subject used reconstructive methods to correct errors, meaning that instead of undoing actions, they simply performed those actions again. For example, to undo an incorrect line spacing, subjects inserted a new line spacing code instead of removing the old, incorrect code. Although such methods will often be effective in working with WordPerfect, they are less likely to be applied on the corrective knowledge test. Consequently, corrective methods were used on this test, and MM+ users were better trained in using these methods for syntactic errors. The issue of how subjects correct their own errors should be addressed in future studies.

There are several reasons why the other expected findings failed to appear. First, the scores on the corrective skill test point at a ceiling-effect. Although the data show a significant difference between the two groups with regard to the detection of semantic errors, their true magnitude cannot be established. For detection, this ceiling-effect may be caused by system cues or the word processor's help function. These built-in resources may have biased the number of errors detected. The results from the corrective skill test can therefore not be seen as an adequate reflection of the actual number of detections.

Second, although preliminary checks on random allocation of subjects to conditions indicated that both groups were identical with respect to prior experience with computers, within-group differences existed. These differences may have affected the assessment of recovery skill. Because more experienced users have a richer, more elaborate conceptual model than less experienced users, their conceptual model allows for a better, more meaningful incorporation of new information. Consequently, they are assumed to benefit more from error information. Future research should therefore focus on how prior experience with computers affects the development of recovery skill.

Third, as with the constructive skill, the actual use of error information may have affected the corrective skill. In cases where error information is not used or explored, MM+ users are not better trained to detect and correct an error than MM− users. Because this experiment revealed no information on how subjects dealt with errors and error information during practice, its true effect on
subjects' constructive and corrective skills cannot be established. In future research on error information, quantitative results should therefore be supported by (observational) data regarding subjects' activity during practice.

Another question for future research is the effect of the cues and prompts generated by the software. During practice, the effect of error information might have been overshadowed by the effect of the cues of the word processor. In future studies, this effect can be eliminated by removing all system cues or by counting the number of times a subject uses this information. Such experiments require an experimental setting that differs from the one that was used here. Individual subjects should be observed during practice as well as during the tests. Not only do these observations provide information on the use of system cues during practice and on the tests, they also reveal more about the actual use of the error information.

The third hypothesis regarding subjects' self-confidence was partly supported by the results. The error information in the MM+ did not cause users to develop higher self-confidence; confidence scores for MM+ users remained rather constant. In the MM− group, however, self-confidence scores after the tests were significantly lower than the scores after practice. So, although the expected increase in self-confidence failed to occur, error information did have a positive effect on self-confidence.

The reasons for the MM+ users' self-confidence to remain constant rather than increase might be that subjects did not know in advance that manual usage was not allowed during the test phase. Intermittent confidence scores might therefore reflect subjects' self-confidence in word processing with the use of a manual. Confidence scores after the tests thus represent self-confidence without the use of the manual. Since most users had little or no experience with WordPerfect, the absence of a manual could have lowered their final confidence score.

The present study showed error information to have hardly any effect on procedural skill (both constructive and corrective). From these findings, one might conclude that including error information in a manual only yields an increase in amount of written information (something manuals can do without!). This conclusion is premature, however. Although error information had no effect on learning outcomes, it could have affected learning activity. In other words, it could have supported users in their development of procedural skills. As the effect of error information on users' activity during practice is still unknown, it can not be decided yet whether including error information in a (minimal) manual is indeed functional.

Acknowledgements

The authors wish to thank Matthias Rauterberg and two anonymous reviewers for their valuable comments on the first draft of this paper.

Correspondence and requests for reprints should be sent to Ard Lazonder, Department of Instructional Technology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands, e-mail: lazonder@edte.utwente.nl.

Provisional findings of this experiment were presented at the 2nd EARLI-SIG

References


Lazonder, A.W. and Van der Meij, H. (in press) 'The minimal manual: is less really more?' Int. J. Man–Mach Studies


Scharer, L.L. (1983) 'User training: less is more' Datamation 29, 7, 175–182


Interacting with Computers vol 6 no 1 (1994)