In the initial experiment, 98 subjects were asked to complete one or two government forms while thinking aloud. Data showed that many problems could be explained by the subjects' tendencies to fail to orientate themselves and to confine themselves to a minimum of effort. Problems were also caused by a lack of relevant background knowledge and of language skills. Based on this understanding of form-fillers' behavior, three general design principles were formulated: forms should be action-oriented; they should control form-fillers' behavior very strictly, and background information should be given on a local level. Seven forms were revised according to these principles. In the second experiment, 86 subjects were asked to fill out one or two of these revised forms, again while thinking aloud. The percentage of forms completed correctly increased from 12.3% to 52.25%. In the third experiment, one form was revised according to the same principles. This revision not only led to an increased proportion of correctly completed forms, but the revised forms were also returned more quickly and were easier to process.

1 Introduction

In one of her trend-setting articles on form research, Patricia Wright (1981, p. 154) states that research on forms must meet two essential criteria:

- The research findings must have practical consequences for the design of forms, i.e. they must be applicable.
- The research findings must generalize beyond the specific design problems of the particular form(s) which motivated the investigation.

One type of possible research findings might be a list of specific guidelines or a checklist. That seems to be the sort of result organizations and designers want. But, as Wright (1981, p.155-157) points out, the guideline approach has significant disadvantages:

- The number of guidelines may grow exponentially so that the list would be unpractical.
- Most guidelines will have exceptions or will be utilizable only in specific circumstances.

A more sensible strategy of research that may produce general as well as utilizable results seems to be a study of form-fillers' behavior, and especially of the problems they encounter filling out a form. When we come to an understanding of these problems, it will become possible to find design principles that might help over-
come them.

The kind of principles we have in mind are of a more abstract nature than simple guidelines. Such principles give credit to the intuition, creativity and skills of designers, and they take into account the fact that certain constraints will be relevant in every individual situation.

Two interesting and important examples of this sort of general design principle in the document design literature of the past decade are the \textit{scenario principle}, formulated by Flower et al. (1983), and the principle of \textit{minimalism} for tutorial computer manuals, advocated by Carroll (1990). In both cases, these principles were expressed in a set of specific text features. Both principles are the result of research into the behavior and the problems of document users.

In this article, we will briefly describe some investigations aimed at attaining a deeper understanding of the behavior and problems of form-fillers. From this understanding, we will derive three basic principles for form design. We will discuss some approaches to a salient design problem: How should designers deal with adjunct information and notes which have to be added to the questions on a form? After reporting some effects of revising forms according to the three principles, we will discuss some limitations of this type of research.

2 Problems of Form-fillers

In 1988 we conducted a large-scale study to gain more insight into practical problems experienced by users of instructional documents (Jansen & Steehouder, 1989; forthcoming). We examined nine government forms. Three of them were from the tax authorities, five from the Ministry of Education and Science, and one from the city council of Hengelo. We asked 98 subjects to complete one or two of these forms on the basis of a situation description, which was intended to put them in the position of a fictitious individual with his or her particular financial situation. We asked our subjects to think aloud while performing their task. Afterwards, the answers were discussed and the subjects were invited to again comment on the problems they encountered. The problems the subjects reported were recorded and coded by trained observers; the same observers made an abstract of the interviews. Of course, the completed forms were also analyzed.

Since we were not primarily interested in the course of the process per se but rather in the problems arising during task performance, we limited our analyses to the moments when:

\begin{itemize}
  \item subjects made a mistake (\textit{problem of effectiveness})
  \item subjects performed an unnecessary action (\textit{problem of efficiency})
\end{itemize}
• subjects showed a lack of understanding of a regulation (problem of understanding)
• subjects asked the experimenter for assistance (problem of autonomy).

We found some substantial differences between the subjects’ actual behavior and the behavior expected or required by the documents. In short, we came to four general conclusions:
• Form-fillers tend to start working towards their goal without proper preparation. Likewise, they almost never reread their answers on the form and never check their calculations. It is obvious that most form-fillers do not use a deliberate strategy. Rather, the predominant strategy can be characterized as a kick-and-rush approach.1
• Form-fillers restrict themselves to a minimum of what they consider indispensable to reach their goal and skip everything else. At first glance, this strategy seems very efficient. However, explanatory notes are skipped, causing many errors due to form-fillers’ missing essential information. In this case, efficiency confronts effectiveness.
• Form-fillers have insufficient background knowledge to adequately interpret the questions and explanatory texts, and to make the right decisions. They hardly seem to realize that this background knowledge is necessary to understand the questions. But even if they do realize, in many cases the chance of gaining this information by reading explanations is neglected.
• Many form-fillers lack language skills and familiarity with graphic signals required by the document.

These conclusions seem in line with the results of several comparable studies performed on a smaller scale in other countries. For instance in a thinking-aloud experiment, Frohlich (1986) asked eight subjects to fill out a form to apply for a Supplementary Benefit. Frohlich analyzed the routes the subjects followed through the forms and the mistakes they made following these routes and answering the questions. He concludes that the standard mode of form completion is to answer the questions in the order indicated on the form and with minimum effort until explicitly instructed otherwise, or until an apparently irrelevant question is encountered. According to Frohlich, the form-filler is easily tempted to overlook important explanations or routing instructions. The result is that he or she provides incorrect answers to relevant questions and fails to skip irrelevant questions.

Frohlich’s findings are quite similar to ours. He concludes that his subjects focused so much on what they considered their real task (answering the questions) that they tended to neglect the required preceding activities. Other small-scale studies into form-filling behavior, for instance a study by Holland & Redish (1981) and one by the German researcher Lüdenbach (1984), sketch the same picture.
3 Design principles

3.1 General principles

We formulated three general principles for the improvement of government forms based on our understanding of form-fillers' behavior: the action principle, the direction principle, and the local background principle.

*The Action Principle*

The first principle reflects the view that completing a form requires a series of actions and that the form can be seen as a *tool* that helps the form-filler perform this task. Consequently, attention should not be focused on general juridical rules, nor upon definitions of terms, but on the actions the form-filler has to perform to complete the form. The action principle resembles the previously mentioned scenario-principle and Kern et al.'s *performance orientation* principle (1975).

*The Direction Principle*

Given the tendency of form-fillers to use a *kick-and-rush* strategy, the direction principle seems appropriate. Form-fillers should be prompted to perform the actions as accurately as possible, in strict conformity with the directions.

*The Local Background Principle*

Form-fillers are not interested in general background information on the government regulation per se; they only need specific information to answer the questions correctly. Therefore, we formulated this third design principle. Explanatory background information on regulations and procedures should be supplied, but the selection and organization of background information should always be geared towards adequate performance of the required actions.

3.2 Example: How to select, formulate and position adjunct information?

Obviously, the three principles can be specified by a number of concrete text features. Here we will confine ourselves to an analysis of the consequences of the three principles for one of the many design problems related to forms.

In addition to the questions on a form, a variety of other information is often provided. The negative attitude most form-fillers have towards this adjunct information is illustrated by the negative connotations of the phrase *the small print*. In our study of form-fillers behavior it became clear that form-fillers generally prefer *not to read* those passages. For instance, despite the clear command at the
top of almost every form to read the explanations carefully before completing the form, only a few subjects did so. And even those subjects stopped reading the explanations after two or three paragraphs. "This is only information," one of them stated.

An analysis of the forms used in the experiment showed eight types of adjunct information:

- **General background information**, summarizing the regulation and related government policy.
- **Functional explanations**, clarifying the purpose of the form and the procedure of which the form is a part.
- **Coding instructions**, indicating how to fill out the form (thick, underline, put answers in boxes, etc.).
- **Definitions**, explaining the legal terms used in the questions.
- **Reasons**, explaining why particular data is needed.
- **Routing instructions**, helping form-fillers to skip irrelevant questions ('go to') or directing the form-filler to relevant definitions of instructions.
- **Instructions on annexes**, making clear under what circumstances annexes have to be sent with the form and what kind of annexes are needed.
- **Information on outcomes**, enabling form-fillers to predict the outcome of the procedure (for instance, the sum of the tax assessment).

From our three principles some specific suggestions can be derived regarding the design of adjunct information.

- The **action principle** suggests that adjuncts that are not directly connected to the actions of the form-filler should be marked as being of less importance. This includes the categories of general background information, functional information, reasons, and information about outcomes. Such text elements should be placed separately from questions and other explanations, since reading them is optional.
- The **action principle** implies that **definitions** should generally be presented not as descriptions, but as **instructions**. For instance, to clarify the concept **saleable value of your house**, instructions should be given (How to assess the saleable value?), instead of clarifying the meaning (What is meant by saleable?).
- The **direction principle** seems of special importance for routing instructions. Clear and explicit routing instructions should guide the routing of the form-filler as strictly as possible.
- The **local background** principle suggests that adjunct information should be given in short portions of texts, located close to the question to which they apply. One option is to use what we have called the **three-column approach**, as demonstrated in figure 1. A form is divided into three columns:
• a left-hand column containing definitions and background information;
• a central column containing the questions themselves;
• a right-hand column containing the answer space and boxes, as well as route instructions (go to).

<table>
<thead>
<tr>
<th>Explanatory notes</th>
<th>Questions</th>
<th>Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>If you are a married woman, only fill out your maiden name here.</td>
<td>1 Your surname:</td>
<td>……………………</td>
</tr>
<tr>
<td>If you are married but not living together with your spouse, tick the no box.</td>
<td>2 Your initials:</td>
<td>……………………</td>
</tr>
<tr>
<td>3 Are you married?</td>
<td>□ yes</td>
<td>□ no</td>
</tr>
</tbody>
</table>

Figure 1. An example of the three-column approach

4 Study of the Effects of Design Principles

Are our principles really practical and do they lead to improvements in forms? In order to get an answer to these two questions, we applied the three principles in revising seven forms and tested the redesigned versions. Again we asked 86 subjects to complete one or two forms while thinking aloud. The same case descriptions were used. Once again we asked the subjects to think aloud and again their remarks were tape-recorded.

The proportion of correctly completed forms increased from 12.3% (original forms; N=106) to 52.2% (revised forms; N=109). This is clearly a considerable improvement, but in absolute terms the results were still a bit disappointing. One point should be noted, though, in connection with these figures. All forms, original and new, were thoroughly tested. We arranged the cases in such a way that the subjects stood a good chance of encountering problems. It was hardly surprising then that these problems actually arose. As a result, however, no direct conclusions could be drawn about the precise percentage of forms that would in actuality have been properly completed. Nonetheless, in view of the extent of the difference, it seems likely that similar improvement would be achieved if the new forms were used in real life situations.

Two years after these experiments, the principles and the presentation techniques following from them were investigated in a more authentic project (Jansen, Klatter & De Vet, 1991). The 'Informatiseringsbank', a division of the Dutch Department of Education and Science, asked for assistance in improving a specific form that
caused them serious processing problems. This form, called 'the parent form', has to be completed by many Dutch parents of students applying for an educational grant. Every year since this form was adopted, the 'Informatiseringsbank' has had to return some 60,000 copies to respondents because of incorrect or missing answers.

In this study forty parents of students were asked to complete the form while thinking-aloud and comment on it afterwards. Twenty of the subjects used a case description, while the other 20 completed the form for their own situation. The findings illustrate the same problems as in our first study.

- Approximately 30% did not read the accompanying letter. Of the remaining 70%, 40% rated it as poorly written.
- Approximately 30% of the subjects did not read the explanatory notes provided on a separate sheet.
- Approximately 25% of the subjects who had to calculate their net annual income produced the wrong figure.

The subjects were also asked to compare the parent form with other forms with which they were familiar. Approximately 25% considered the parent form 'more difficult', approximately 30% considered it 'easier' and approximately 40% did not notice any difference.

In the following stage, two new versions of the form were designed using the principles stated above. The chief difference between these two new versions is that one (form P) was made with the same rather outdated laser printer hardware used for the original version of the form, while the other (form Q) was printed on a more sophisticated laser printer.

<table>
<thead>
<tr>
<th></th>
<th>Original (N=40)</th>
<th>Revised P (N=40)</th>
<th>Revised Q (N=40)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All questions completed correctly</td>
<td>5</td>
<td>20</td>
<td>15</td>
</tr>
<tr>
<td>Forms rated as more difficult than average</td>
<td>24</td>
<td>15</td>
<td>5</td>
</tr>
<tr>
<td>Accompanying letter rated as poorly written</td>
<td>40</td>
<td>20</td>
<td>5</td>
</tr>
</tbody>
</table>

*Table 1. Effects of revising the parents form of the Informatiseringsbank*

Just as with the original version, both new forms were filled out and commented on by forty subjects. Table 1 shows that both new versions were completed correctly more often than the original form. The new forms were also rated more positively. Furthermore, form Q, made with the more sophisticated laser printer,
was rated somewhat higher than form P. The same applies to the accompanying letters. It must be stressed that the wording and the overall structure of both were identical. It seems that just changing letter type and printing density influences people’s assessment of the difficulty of forms and letters. Despite these differences in assessments, actual performance on both revised forms did not significantly differ. This suggests that users’ assessments may not always be a valid measure of usability.

In order to get an indication of how the organization would actually profit from redesigning the forms, two samples of one hundred new forms were sent to parents of students all over the country. These parents were not told that the form they received was in an experimental stage. They were merely asked to complete it. Another group of one hundred parents received the original form, and they had to perform the same task. After a period of six weeks, the numbers of forms that had been sent in were counted and the parents’ answers were analyzed.

It turned out that both revised forms were returned to the Informatiseringsbank more quickly than the original. It also happened that, while using the original form leads to a yearly figure of about 60,000 copies that cannot be immediately processed and have to be returned, introduction of form P or form Q would decrease this to some 15,000 or 20,000 copies. This results in a reduction from some 12% to 3% or 4%.

An interesting issue concerns the influence of the experimental setting upon the performance of subjects, in particular the influence of using case descriptions. As table 2 shows, overall performance with case descriptions was considerably poorer than performance when filling out a form according to one’s own situation. This applies to the original as well as to the revised forms.

<table>
<thead>
<tr>
<th></th>
<th>original</th>
<th>revised</th>
<th>revised</th>
<th>revised</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>P</td>
<td>Q</td>
</tr>
<tr>
<td>Own situation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Completed correctly</td>
<td>8%</td>
<td>33%</td>
<td>42%</td>
<td>28%</td>
</tr>
<tr>
<td>Case description</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Completed correctly</td>
<td>4%</td>
<td>15%</td>
<td>4%</td>
<td>8%</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Completed correctly</td>
<td>5%</td>
<td>20%</td>
<td>15%</td>
<td>13%</td>
</tr>
</tbody>
</table>

Table 2. Some effects of using case descriptions or subjects own situation
5 Discussion

Some reservations should be mentioned about the research sketched in this article. For instance, the working-aloud method of problem research as we have applied it is not unquestioned. In addition, the use of case descriptions has undoubtedly influenced the behavior and the problems of the form-fillers.

We will not discuss these reservations in detail here but will confine ourselves to the issue of research strategy. First of all: Is the study and analysis of user problems really a straightforward method for establishing generally practicable design principles? And secondly: Is revising a complete document and measuring user performance a proper method for validating those principles?

6.1 Other Methods of Establishing Design Principles

The present research provides a good perspective: It proved possible to establish and support design principles based on an understanding of user problems. However, not all form problems will to come to light when the emphasis is placed on effectiveness and efficiency from the viewpoint of the form-filler. In this respect, previous research (not only ours) has been limited in two ways.

Firstly, there are at least two addressees for every form: the form-filler and the employee who has to process the form. Studies starting from the viewpoint of those responsible for processing the forms are very rare. Perhaps the only exception is the studies of Barnard, Wright and Wilcox (1976, 1978) on the legibility of spaced character formats. Although the parent form study showed that optimizing usability from the form-fillers perspective may also increase processing effectiveness and efficiency, not much attention has been paid to this aspect in research.

Secondly, forms can serve more goals than just transferring information effectively and efficiently. Studies related to other form functions are also very scarce. As far as we know, no systematic analysis of functions, other than from the perspective of the form-filler, has yet been published. Nevertheless, a brief examination suggests divergent functions such as

- establishing goodwill for an organization
- preventing fraud
- encouraging citizens to apply (or perhaps discouraging them from applying) for a particular benefit.

It seems very likely that more thorough analysis would produce many more functions that would result in more, and possibly conflicting requirements, of form design.
But even if we confine ourselves to the perspective of the form-filler, research on form filling behavior will not be sufficient to detect all problems and provide a basis for all design decisions. One topic that is very unlikely to be detected in a usability test is the order of questions on a form.

An example may illustrate this point. In figure 4 there are two questions to determine whether an applicant is entitled to a particular tax refund. It seems reasonable to suppose that question 2 (Do you expect your yearly income in 1992 to exceed that of 1991 by 25% or more?) is much easier to answer than question 1 (Is the average of your yearly income from 1986 up to 1990 more that 125% of your income in 1991?). Therefore, it may be advantageous to change the order and ask question 2 first. In that case, many form-fillers could skip the difficult question and go directly to 3.

However, when the percentage of form-fillers who expect their income in 1992 to exceed that of 1991 by more than 25% is extremely small, the original order might be more efficient, because it prevents the vast majority from processing question number 2 which is irrelevant to them.

In general, the most efficient order of questions depends on two factors: the (estimated) proportion of form-fillers who will answer each question with yes \(p\) and the (estimated) time form-fillers need for answering each question. If both questions \(A\) and \(B\) must be answered with yes to reach a certain outcome (in this example: question 3), the underlying rule can be formalized as a conjunction: \(A \& B \leftrightarrow P\). In this case, there are two possibilities: either \(A\) first and \(B\) second \((A, B)\), or the opposite \((B, A)\).

Which of these possibilities is the most efficient? To decide, we need to know the probability that question \(A\) will be answered positively \((pA)\); we need to know the possibility of a positive answer for \(B\) \((pB)\), and we need to know the average amount of time required to answer question \(A\) and question \(B\), respectively \((tA and \(tB)\).
When the values of \( p_A, p_B, t_A \) and \( t_B \) are known, the average time readers will need to follow each of the two possible courses of actions can be calculated.

If the order of the questions is \((A, B)\), all form-fillers have to answer question \( A \); only a portion of them \((p_A)\) also have to answer \( B \). Thus, the average total amount of time required to reach the outcome will be: \( T(A, B) = t_A + p_A t_B \). If the order is \((B, A)\), the average total amount of time will be: \( T(B) = t_B + p_B t_A \).

Of course, the most efficient possible course of actions will be the one where \( T(X, Y) \) is least. Mathematically, the order \((A, B)\) is most efficient if and only if:

\[
\begin{align*}
  t(A, B) &< t(B, A) \\
  \Rightarrow tA + pA * tB &< tB + pB * tA \\
  \Rightarrow tA - pB * tA &< tB - pA * tB \\
  \Rightarrow \frac{tA - pB * tA}{tA * tB} &< \frac{tB - pA * tB}{tA * tB} (tA > 0 \text{ and } tB > 0) \\
  \Rightarrow \frac{1 - pB}{tB} &< \frac{1 - pA}{tA} \\
  \Rightarrow \frac{1 - pA}{tA} &> \frac{1 - pB}{tB}
\end{align*}
\]

In other words, given a (part of a) form where two questions have to be answered with yes to ensure a specific outcome, the most efficient question order begins with the question for which it is true that \( \frac{1 - p}{t} \) is maximum.

In the same way, it can be shown that, given a (part of a) form where one of two questions have to be answered with yes \((A \lor B \Leftrightarrow P)\), the most efficient order begins with the question for which it is true that \( \frac{P}{t} \) is maximum.

It is very unlikely that even in a simple case like figure 4, an inefficient order of questions will be detected by form-fillers in the kind of experiments described in this article. Only logical analysis of text structure will detect such problems. In another study we analyzed the criteria for an optimal order of questions in greater detail, and we developed a heuristic for designing such an optimal order (Steehouder & Jansen 1991).
Validating General Design Principles

We began this paper arguing that general principles rather than specific guidelines are of the greatest importance for document design. They take into consideration the fact that every document has its own requirements, possibilities and constraints, and that most communication problems are too complex to solve with a limited set of strict guidelines.

However, from a research point of view, this approach leads to considerable difficulties in building and testing hypotheses. When a form is revised according to general principles, many text features are simultaneously manipulated. When differences in performance or in rating are found, it is not exactly clear precisely which text variables caused the effect.

This methodological problem seems far from a solution. Validation of general principles might not even be possible when an attempt is made to satisfy all experimental research demands. But then again, this problem is common to most fields of applied research.

Notes

1 In the American literature, the expression jumping the gun has been used to characterize this behavior. The choice of analogy seems to reflect an interesting cultural difference between the continents.

2 In Jansen & Steehouder (forthcoming) a model is proposed that describes different types of actions involved in completing forms. The model makes a distinction between actions on a functional level (generating information, verifying conditions, transforming data and coding), actions related to interpreting text (semantically and pragmatically), and actions on a monitoring level (orientation, selection, checking and switching between interpreting and functional tasks).

3 It should be emphasized that this behavior is generally not a consequence of inaccuracy, but the result of form-fillers' deliberate decisions (cf. Wright 1989).

4 Examples of revisions are given in Jansen & Steehouder (forthcoming).

5 Forms were considered to be completed correctly if the answers would lead to the correct decision in the given case; this does not necessarily imply that all answers were correct.

6 In this experiment, a form was considered correctly completed only if all questions to be answered were answered correctly and all irrelevant questions remained unanswered. This criterion is thus much stricter than in the previous experiment.
References


About the authors

Michaël Steehouder and Carel Jansen worked together at Twente University from 1976 till 1989. Their research activities included design issues of forms and instructional texts, and the teaching of writing in higher education. They were co-authors of a textbook on communication (*leren communiceren*, 3rd ed. 1992) and of a textbook on forms design (*Formulierenwijzer*, 1989), and together they published *Taalverkeersproblemen tussen overheid en burger* (1989). More recently their research and educational activities extended into the field of user documentation and other instructional texts. Since 1989, Carel Jansen is an Associate Professor at Utrecht University, while Michaël Steehouder is an Associate Professor at the University of Twente.