

Redesigning and testing a work order form

Davida Charney
Carnegie-Mellon University

An order form used in a university printing office was evaluated, redesigned, and tested to determine whether the original or the revised version could be completed more quickly and accurately. Subjects filled out one version of the form for four hypothetical printing tasks. Half the subjects had previous experience using the original version and half did not. The performance of both groups was better on the new version, as measured by the time they needed to complete the forms and the nature of the errors they committed. An analysis of thinking-aloud protocols is presented to explain the apparent superiority of the re-designed form.

It is useful to think of forms as mediating between two groups: the people who fill them out, and the people who read and act on them.¹ Ideally, forms help a group of form-fillers to communicate information that is needed by a separate group of form-readers. One common type of form is the work order, which is a set of instructions from a client who commissions a job to the employees of a business where the work will be performed. For example, the occupants of an office building may need to fill out a maintenance work order to have an office painted or some electrical wiring installed. Work orders present a special challenge to the form designer, because the form-fillers and the form-readers often speak very different languages. The client decides what is to be done on the job, but may have little or no idea what it is that the employees do. The client may want to improve the lighting in the office, but not know what such a job would involve. The employees, however, need decisions to be made about the very points with which the client is unfamiliar: exactly which operations to perform and what materials to use. The work order is supposed to bridge the gap by presenting an array of available options and

prompting for the information that the employees need.

The designer of a work order must keep both the business and its clients in mind. The content of the form is primarily determined by what information the employees of the business need to carry out their work. In order for the form to be successful, however, the clients must be able to fit their wishes to the options provided and encode the required information on the form. Clients should not have to struggle to translate their instructions into the specialised language of the trade. Concern for the client, then, should guide the spatial arrangement of the items, the language used in the questions and the kind of response that is required. Wright and Barnard (1975, 1978) and others (for example, Lefrere et al, 1983 and Matthews et al, 1982) have demonstrated that these factors greatly affect a user's ability to fill out a form.

This paper reports on a project in which I worked with both the employees and the clients of a printing business to evaluate and redesign a work order and to test the usability of the redesigned version of the form against the original. The form is the basic work order used at Carnegie-Mellon

Davida Charney is completing her doctorate in rhetoric at Carnegie-Mellon University.

Author's address:
Department of English,
Carnegie-Mellon University,
Pittsburgh, PA 15213, USA

© Davida Charney 1984

Information design journal,
4/2, 00-00

REQUISITION PUBLICATIONS/CAMPUS PRINTING Carnegie-Mellon University

Name _____ Department _____ Ext. _____ Job No. _____

Description _____ Date Submitted _____

Upon Completion: Deliver To _____ Will Pickup (Call _____ when ready) Date Requested _____

Hold for Mailing/Extras To _____ Approved _____

SERVICES REQUESTED: Design Composition Key Art

DUPLICATING

Indicate Only If Not: 8 1/2 x 11 8 x 11 Black

Art/Original Must Be Marked for One or Two Sides, Bleeds, Halftone Positions, Second Color, Etc.

Item	No. of Originals (Flats)	Quantity Desired	No. Pages One Side	No. Pages Two Sides	Original Size	Finished Size	Color of Ink	Bleeds	Negative & Plate	Rerun	Half-tones No.	MIF Required	AIS Attached
A.													
B.													
C.													
D.													

Stock (If other than 8 1/2 x 11, 20# Bond, White used)	Date Order	Quantity	Cost	Operator Initials	Date Printed	No. Used	Est.
A.							
B.							
C.							
D.							

CAMERA If Enlarged/Reduced to Particular Dimension, Art Must Be Clearly Marked in Blue Pencil. Xerox Reductions Requested AIS Required

Item	Type of Print				Size of Print		Type/Shot		Size		Final Size % of Original	No. of Prints
	Kodolith Negative	PMT	Reverse Negative	PMT	10x12	12x18	Line	Half Tone	100%	Enlarge Reduce		
I.												
II.												
III.												
IV.												

BINDERY - Indicate each operation with letter code used above. AIS Required

Collate Items _____

STITCHING

Upper Left	Left Margin			Saddle Stitch			GBC	Tape	Velo	Perfect
	1	2	3	2	3					

FOLDING (Submit Sample)

1/2	1/3L	1/3R	1/2x1/2	1/2RL	1/3L	1/3R	Other

OTHER (Submit Sample)

Regular Pad	NCR Pad	Trim	Final Trim	Perforate	Number	Die Cut	Drill Holes
no. pad	no. set	Book	Size	<input type="checkbox"/> See Art	From To	<input type="checkbox"/> See no. Art	<input type="checkbox"/> XL

ADDRESSING/MAILING (Fill out mailing specification form)

DELIVERY DATA:

	Date	Quantity	Initials
Partial			
Partial			

1/78 PUB/CPO 2

STATEMENT OF CHARGES

Design (1-08120-7816) = \$ _____

Composition (1-08120-7814) = _____

Key Art (1-08120-7817) = _____

Offset (1-08120-7811) = _____

Copier (1-08120-7812) = _____

Mailroom (1-08120-7813) = _____

ID Cards (1-08120-7815) = _____

Miscellaneous (1-08120-7780) = _____

1st Class Postage (1-00000-1639) = _____

3rd Class Postage (1-00000-1635) = _____

TOTAL CHARGES \$ _____

Authorized By _____

Budget No. _____

Samples/Art to Publications _____

Art Originals to Customer _____

Figure 1

University's Campus Printing Office (hereafter referred to as the 'CPO'). The CPO provides the university community with various services, ranging from type-setting and camera work to photocopying, off-set printing and binding. To request one or more of these services, clients complete and submit the 'Requisition' pictured in Figure 1. The clients of the CPO, including some longtime customers, often had difficulty filling out this form. As a result, one CPO employee spent all her time helping clients fill out the forms, inspecting forms which were sent in by mail, and telephoning clients for clarifications. Efforts to improve the situation could not centre on educating the client; it would be impractical to attempt to teach the client the trade or even a portion of its terminology. However, it is possible to help clients infer what they need to know by ensuring that the form presents them with clear alternatives in terminology they can understand.

In the summer of 1982, the director of the Office of Publications and Campus Printing hired Carnegie-Mellon's Communications Design Center to evaluate all the forms used in the CPO. Eventually we identified five forms as needing major revision, but from the outset it seemed clear that our attention should focus first on the Requisition. The Requisition was the basic channel of information from the client to the CPO, and the operations it described involved every CPO department.

Evaluating the requisition

The first step in evaluating the form was to review its content in order to learn how to fill out the form correctly, to identify trouble spots on the form, and to anticipate the consequences of eliminating or changing particular boxes.² Three methods were employed in the review. First, interviews were

conducted with those CPO employees who had to extract information from the form in order to carry out their jobs. These interviews were designed to elicit the following kinds of information:

- what the terms on the form meant, and whether these definitions were consistent throughout the CPO;
- what information each box was intended to elicit;
- to what extent the entries for any given box could vary, and what kinds of entries were most typical;
- which boxes clients filled out incorrectly most often;
- which sections of the form were out of date, for example, offering services which were no longer available, or neglecting to offer new services;
- what path the form normally took through the various departments of the CPO, as a job proceeded.

In addition to this information, it was important to get a feel for what kinds of services the CPO's clients requested most frequently. Ideally, the Requisition should accommodate the typical jobs most easily, and unusual requests secondarily. For example, the Requisition has lines for four separate items in the Duplicating section. If a large proportion of jobs required five or more items, it might be necessary to expand this section. In order to learn what kinds of jobs were typical, a survey was conducted of the backfiles of jobs from the previous year.

Finally, the form was analysed for the kinds of problems discussed in Wright and Barnard (1975): unclear language, illogical sequence of questions, insufficient space, and so on. In the end, three major categories of problems were identified in the present version of the Requisition.

Function

Although the major purpose of the Requisition was to record the client's instructions for a job, this function was obscured by the inclusion of items intended for other purposes. The grey areas on the form were intended to help the CPO staff keep track of which operations had been completed as the job proceeded through the shop. For example, employees in the Bindery were supposed to record what they had done on the job in the grey box below the appropriate Bindery option (see Figure 1). The Requisition also served a third purpose: when the job was finished, the 'Statement of Charges' section was filled out, and a copy of the form was sent to the client as an invoice.

The form functioned least well as a means of keeping track of the work completed on the job. The grey areas are too small to contain the necessary information, especially in the Camera and Bindery sections. Most CPO employees actually recorded their work on a separate page. Since the grey areas were not being used effectively, and since they seemed to break up the form visually, their presence detracted from the form's primary function, and they could probably be eliminated without serious consequences.

Language

The language used to label the boxes is often ambiguous, confusing or overly technical. Some of the labels are simply difficult to parse. In the Duplicating section, for example, two labels impose negative conditions: 'Indicate Only if Not', and 'Stock (If other than 8½x11, 20 #, Bond, White used).' Other labels in the same section are ambiguous in reference. The two labels 'No. Pages One side' and 'No. Pages Two Sides' are particularly open to interpretation: do they refer to pages of originals or pages of copies? Finally, a number of labels contain technical terms, abbreviations or trade names that general audiences

would not be familiar with, such as 'bleeds', 'half-tones', 'PMT', 'GBC' and 'saddle stitch'.

Layout

Much of the form is organised into a series of grids or matrices, creating a false impression of uniformity among the options and obscuring the different relationships among them.

One problem with the grid format is that it is often difficult to group options within it. In this form the client is sometimes asked to choose one of a set of mutually exclusive options. At other times, the client may choose several options from a group of compatible options. However, because the groups of options all look alike, it is difficult to distinguish boxes which present either/or choices from those that do not. For example, consider the right hand portion of the Camera section. Under 'Size', the client chooses '100%' or either 'Enlarge' or 'Reduce'. If he chooses one of the latter two, then he must also fill in the next box, 'Final Size % of Original', but this box is not grouped under 'Size' with the others.

A similar problem occurs in the Bindery section of the form, pictured at the bottom of Figure 1. Although they look very similar, the options under 'Stitching' constitute a mutually exclusive set, while the options under 'Other' do not. Furthermore, the grid format conceals the fact that the 'Stitching' options fall into two distinct subcategories. 'Upper Left', 'Left Margin' and 'Saddle Stitch' all describe how pages are stapled, while 'GBC', 'Tape', 'Velo' and 'Perfect' describe different binding materials.

A second consequence of the grid format is that it unnaturally restricts the size of the cells. As a result, the space available for writing in the cells is often unrelated to the amount of space required. For example, in the Camera section, a checkmark is all that is required in any of the boxes under 'Type of Print'. Yet these boxes, which are all fairly large, vary in width because of the difference in the

lengths of the labels: 'PMT' vs 'Negative'. More space is allocated to these boxes than is necessary, while elsewhere, the space allowed is inadequate. Several boxes in the Duplicating section are too small to contain the required information, including the box for describing the items, 'Original Size', 'Finished Size', and 'Color of Ink'.

The arbitrariness of the size of the cells has another adverse consequence. The cells in a grid provide no visual cues for how to respond to a box. Because the cells are all approximately the same size, it is unclear which are supposed to contain checkmarks, and which words or numbers. As a result, the client has no clue apart from the label on a box as to whether the box presents an option or calls for a description.

Redesigning the form

Based on this analysis, the art director for the Office of Publications and I created a new version of the form. Our goal was to create a form that would retain the essential features of the present form, but which would be clear enough for clients to fill out accurately and without assistance (at least for most jobs). Sketches of the new version were shown to each of the CPO employees who had been interviewed, and many of their comments and suggestions were incorporated into the version that was eventually tested. This version, entitled 'Work Order', is shown in Figure 2. A few of the major changes in the form will be discussed here.

Removing the grey areas from the old version of the form provided a fair amount of room to work in. The grids were replaced with two kinds of boxes. One kind calls for a verbal description and consists of a label and a line where the client can write words or numbers. The second kind of box presents the client with an option. This kind of box consists of a label and a checkbox. In general, mutually exclusive options are signalled by conjoining two or more checkbox options with the word 'or'.

WORK ORDER

Date Submitted _____

Client _____ Ext. _____ Date Requested _____

Title of Job _____ Approved By _____

Send Charges to Name _____ Department _____ Date Completed _____

Publication See AIS Writing/Editing Design Photography Composition Art

Duplication See AIS See Original Xerox Reduction (Reduce original to _____)

Item A _____ No. of Originals _____ Quantity Requested _____

Finished Flat Size _____ Single Sided or Duplex Ink _____ Bleed Rerun

Stock _____ Stock Supplied

Item B _____ No. of Originals _____ Quantity Requested _____

Finished Flat Size _____ Single Sided or Duplex Ink _____ Bleed Rerun

Stock _____ Stock Supplied

Item C _____ No. of Originals _____ Quantity Requested _____

Finished Flat Size _____ Single Sided or Duplex Ink _____ Bleed Rerun

Stock _____ Stock Supplied

Camera See AIS See Original

Item 1 No. of Originals _____ No. of Prints _____ Same Size Enlarge or Reduce Original to _____

10 x 12 or 12 x 18 Print Size Line or Halftone PMT or PMT Reverse or Transparency or Film Negative

Item 2 No. of Originals _____ No. of Prints _____ Same Size Enlarge or Reduce Original to _____

10 x 12 or 12 x 18 Print Size Line or Halftone PMT or PMT Reverse or Transparency or Film Negative

Item 3 No. of Originals _____ No. of Prints _____ Same Size Enlarge or Reduce Original to _____

10 x 12 or 12 x 18 Print Size Line or Halftone PMT or PMT Reverse or Transparency or Film Negative

Bindery See AIS See Original See Sample

Collating Collate Items (See AIS) Slip Sheet Insert (See AIS)

Stapling Side Stitch (2 or 3) Saddle Stitch (2 or 3)

Upper Left (One Staple)

Binding Plastic Comb Plastic Strip Perfect

Folding 1/2 1/2 x 1/2 1/3 Z LTR

1/2 RL 1/2 1/2 RL 1/3 LTR Other (See AIS)

Finishing Trim Perforate Number (From _____ To _____)

Die-Cut Drill (No. of Holes _____) NCR

Shrinkwrap (_____ per pkg.) Pad (_____ pads w/ _____ per pad)

Delivery See AIS

Deliver to Bldg. _____ Rm. _____ Call Ext. _____ for Pickup

See Mail Specifications Hold for Mailing (extra to Bldg. _____ Rm. _____)

Authorized Signature _____

Budget Number _____

Charges

Design 1-08120-7818 _____

Photography 1-08120-7819 _____

Writer/Editor 1-08120-7818 _____

Composition 1-08120-7814 _____

Key Art 1-08120-7817 _____

Offset 1-08120-7811 _____

Copier 1-08120-7812 _____

Mail 1-08120-7813 _____

1st Class 1-00000-1639 _____

Permit 1-00000-1636 _____

ID Cards 1-08120-7815 _____

Miscellaneous 1-08120-7760 _____

TOTAL CHARGES \$ _____

Figure 2

Eliminating the grids made it possible to create natural subsections. For example, the Duplication section contains subsections for three items, each of which incorporates stock information for that item. This seems an improvement over the old version of the form, in which the stock section appeared below the Duplicating grid, and clients were expected to describe the stock for an item on the appropriately lettered line. Similarly, subdividing the Bindery section was useful for setting off the major operations and clarifying the suboptions.

Finally, general terms were substituted for technical terms wherever possible. For example, the 'Stitching' section was broken into two separate sections with the labels 'Stapling' and 'Binding'. The trade names 'GBC' and 'Velo' were replaced by descriptive labels, 'Plastic Comb', and 'Plastic Strip', respectively. The two boxes 'No. Pages One Side' and 'No. Pages Two Sides' from the Duplicating section were replaced with checkboxes: '[] Single-sided or [] Duplex'. The use of the term 'Duplex' to mean 'printing on both sides of a page' is familiar to our University community, since it is used in the command for printing files from the computer. The name of the form was also changed from 'Requisition', which connoted ordering supplies, to 'Work Order'.

Comparing the old and new forms empirically: an overview

Intuitively, we felt that the new version of the form would be easier for clients to use. However, in order to justify replacing the old version of the form, we first had to demonstrate that the changes would significantly improve a client's ability to fill out the form. We were interested in the effect of the changes on two kinds of clients: regular CPO

clients, many of whom had grown accustomed to the old version of the form, and potential CPO clients, who would have to fill out the form for the first time. The new form should benefit both groups: the regular clients should not have to struggle to learn to use the new form, and potential clients should have an easier time learning what their options are.

We therefore conducted a study in which we observed both current and prospective CPO clients filling out the two forms. We found that, overall, both groups of clients filled out the new version of the form more quickly than the old. The subjects made about the same number of errors on the new form as on the old, but the errors that were committed on the new form were of a less serious nature. In addition to collecting such quantitative evidence, we also collected 'on-line' observational evidence about how subjects went about completing the form, and which parts of the form gave them trouble. This was accomplished by asking subjects to think aloud while filling out the forms, and tape-recording their comments.

Such 'thinking-aloud protocols' are useful in two important ways. First, the protocols supplement the statistical error analysis: they show us what is happening as the subject makes a mistake, aiding the analysis of what caused the problem. Does the error occur because subjects add some numbers incorrectly, or because they misunderstand what information is required? Second, protocols identify trouble spots that might be missed by simply counting errors on the completed form. For example, suppose subjects spend a great deal of time and effort deciding how to fill out a confusing box correctly. If we simply score the form for errors, this box would not appear to be a trouble spot, because it is filled in correctly. However, the protocol can show us that subjects read and re-read the label on the box, and consider several different entries before hitting on the correct one. By

listening to what the subjects are thinking as they fill out the forms and make their mistakes, we are better able to see what causes the problems, and make informed judgements about how to solve them.³

The study provided confirmation that most of the major problems in the old version of the form had been identified, and that the new form contained effective solutions for most of these. The study also revealed some problems in the new form, which led to further revisions of that form.

Method

Design

Two types of subjects were used: 'experienced' subjects who had previously used the old version of the form to order work from the CPO, and 'novice' subjects who had never used the form before. Subjects of both types were randomly assigned to one of two groups: either they filled out the old version of the form or they used the new version. Each subject filled out the appropriate version of the form four times, once for each of four tasks. Performance on the tasks was measured in terms of the time subjects required to complete each task and the number and nature of the errors in the completed forms.

Subjects

The subjects were 12 full-time employees at Carnegie-Mellon University who volunteered to participate in the study without compensation. Six of the subjects were 'experienced', and six were 'novices'. Each member of the experienced group had used the old version of the form more than once to order work from the Campus Printing Office, while none of the novices had ever seen the form before or dealt with the Campus Printing Office in any way.

The experienced subjects were fairly representative of the regular clients of the CPO. For

example, the degree of their familiarity with the form varied; some had only used the form a few times, while others routinely filled it out several times a month. In addition none of the subjects was familiar with all of the services the CPO had to offer. This meant that the experienced subjects were not equally familiar with all four of the tasks they were asked to do.

Materials

Four tasks were prepared which represented the range of services provided by the Campus Printing Office, and which together required subjects to fill out boxes in every section of the form.

- The 'Report Task' called for ordering 250 copies of a technical report and binding the reports in blue covers, using a plastic spiral binding.
- The 'Design Task' called for making an enlarged print of a logo using a photo-mechanical transfer technique.
- The 'Invitations Task' called for printing invitations, reply cards and envelopes. These pieces were of different sizes and had to be printed on different kinds of paper. In addition, the invitations had to be folded in quarters.
- The 'Class Materials Task' called for photocopying an article, making reductions so that each side of the final copy contained two pages of the original.

The tasks were designed to be as natural as possible: giving subjects enough information to fill out the form, but not so much information that the task would be trivial. The instructions for each task set up a hypothetical situation in which the subject acted as a client of the Campus Printing Office, ordering some work. In each task, the subjects were given a set of 'originals' to be submitted to the CPO, a blank form, and some samples of different paper stocks. The paper samples were labelled with a

trade name, weight and colour. This information was necessary for filling out the Duplicating sections of the forms and was supplied to the subjects because they were unlikely to know it on their own. For two of the tasks, subjects were also given a facsimile of the final product. For the Report Task, for example, they were given a copy of a technical report bound in heavy blue paper with a plastic spiral binding. The facsimile served two purposes. First, it increased the realism of the task, since clients of the CPO often bring in examples of what they want the final product to look like. Second, the facsimile made it possible to convey the goal of the task non-verbally, so that the instructions would not give away the actual terms on the forms. For example, the instructions for the Report Task did not mention covers or a binding; subjects inferred that they had to order these by comparing the model end-product to their set of originals.

Procedure

Within each group, subjects were randomly assigned to either the old version or the new version of the form. Each subject was run individually, in sessions which took approximately one hour for the four tasks. First, subjects were given general instructions for giving a thinking-aloud protocol: they were told that the session would be tape-recorded and were asked to think aloud as much as possible while filling out the forms. Subjects were then given the materials for the first task. They were asked to read the instructions for the task aloud and then to fill out the form as if they were the client ordering the work. As an example, here are the instructions for the 'Report Task'.

'Your department occasionally publishes technical reports. You have been put in charge of getting 250 copies of the supplied report printed and bound with covers. The report entitled 'Studies of Metaphor and Complex Analogies' is the report you wish to have copied. You are to fill out the form in such a way that your report

will come out looking just like the technical report entitled 'How Headings in Documents can Mislead Readers'. Make sure the title page, the page of tables and the page of figures are left blank on the back. The reports must be ready for you to pick up next Monday.'

Subjects then filled out the form for that task. As they worked, subjects were allowed to ask the experimenter to define specific terms on the form. After a subject completed the first task, the experimenter took away the materials for that task and supplied the materials for the next one. The experimenter also recorded the starting and ending times for each task. The order in which the four tasks were presented to the subjects was randomised. After completing all four tasks, the subjects were given copies of both versions of the form, and were asked to comment on them.

Scoring procedure

The twelve subjects each completed four forms. Each of these 48 forms was scored for the number of errors the subject made, and the nature of the errors. Subjects made the following three types of errors, which are listed in order from least to most serious:

1. Omission errors. Skipping a required box; that is, leaving blank a box that should have been filled in.
2. Misplaced entries. Marking an irrelevant box; that is, filling in a box that should have been left blank.
3. Wrong entries. Marking a relevant box incorrectly.

The seriousness of each type of error was determined in consultation with the CPO staff. Omission errors were judged to be least serious because they are the easiest for the CPO staff to detect and correct. For example, if a client doesn't fill in the 'Quantity Required', the job cannot be completed, and the client is contacted for the

necessary information. Similarly, if a client indicates that a set of reports should be stapled, the staff can easily infer that the pages must be collated, even if the collate box is left blank.

Misplaced entries and wrong entries are more serious. In either case, the client has marked a box, and the CPO employees justifiably assume that that option was chosen deliberately. Unless the error is detected, the staff will attempt to carry out the instructions as written. Misplaced entries are slightly easier to detect than wrong entries. Marking an irrelevant box often results in instructions that are incompatible with the rest of the job. On the other hand, a wrong entry can be hard to detect because the choice may seem perfectly reasonable, even though it does not reflect what the client knows or wants. At the very least, wrong entries cause confusion. For example, if a client enters the wrong number of originals, the CPO staff will waste time tracking down the 'missing' pages. In the worst cases, these errors lead the CPO to faithfully execute the wrong operation. For example, suppose a client who wants a photograph reduced by one-third writes '30%' instead of '70%' in the 'Final Size % of Original' box. Unless the error is detected, this client will end up with a much smaller print than expected.

Results and discussion

The performance of the subjects was analysed in terms of how much time was required per task, how many errors were committed per task, and what kinds of errors were committed.

Consider first how much time subjects needed to complete each form. The average time per task (and standard deviations) for subjects in the four conditions is summarised in Table 1. Not surprisingly, the experienced subjects were faster at completing the forms than the novices. The experienced subjects took an average of 10.7 minutes to fill out each form, while the novices

	Old form	New form	
Novice	1.00 (3.78)	13.33 (3.00)	14.67
Expert	13.00 (2.84)	8.33 (1.42)	10.67
	14.50	10.82	

Table 1. Average time per task in minutes, and standard deviations, for subjects (n = 12) in the four conditions.

required an average of 14.7 minutes. A two-way analysis of variance on the version and experience factors showed this speed advantage to be statistically significant, $F(1,8)=5.75$, $p < .05$. Table 1 also indicates that overall, subjects were able to fill out the new version of the form more quickly than the old version. The average time for subjects to fill out the new version of the form was 10.8 minutes and for the old version, 14.5 minutes. This advantage was marginally significant, $F(1,8)=4.83$, $.05 < p < .06$. The marginal result seems due to the smallness of the sample size (three subjects per condition), rather than to the size of the advantage.

Next consider the number and kind of errors that appeared on the forms. The results show that subjects working on the two versions of the form made approximately the same number of errors overall. However, the errors that subjects committed on the new version were less serious in nature than the errors on the old version.

Table 2 shows the average number of errors (and standard deviations) for the subjects in the four conditions. This measure produced no main effect of either experience level or version. That is, although the novices made a few more errors than the experienced subjects, and subjects made a few more errors on the old version than on the new, these differences were not statistically reliable. However, there was a significant interaction between version of form and level of experience. Novices tended to commit more errors on the new version of the form (8.75 errors on the old form as compared to 10.66 on the new), while the experienced subjects made more errors on the old version (9.83 errors on the old form as compared to 6.41 on the new). A two-way analysis of variance on the experience and version factors indicated that this interaction is significant, $F(1,8)=5.35$, $p < .05$. The fact that the experienced subjects' performance tended to improve when they used the new form is suggestive. One might expect that the

	Old form	New form	
Novice	8.75 (2.70)	10.66 (2.43)	9.70
Expert	9.83 (1.61)	6.41 (0.38)	8.12
	9.29	8.23	

Table 2. Average number of errors per task, and standard deviations, for subjects (n = 12) in the four conditions.

		Type 1	Type 2	Type 3
Novice	Old version	2.637	3.33	8.33
	New version	33.30	6.33	3.00
Expert	Old version	25.67	4.00	9.66
	New version	19.00	3.33	3.67

Table 3. Average number of errors per task, by type of error, for subjects ($n = 12$) in the four conditions. Type 1 = omission errors; type 2 = misplaced entries; type 3 = wrong entries.

experienced subjects would have some trouble making the transition from the version they were familiar with to the new version of the form, but instead it seems that their experience with the old form facilitated their efforts with the new form. It seems discouraging that novices made more errors on the new version than the old. However, this result appears less serious when the nature of the errors is analysed.

An analysis of the types of errors subjects made suggests that while the total number of errors committed on the old and new versions of the form is the same, the errors committed on the new form are less serious. In particular, when subjects used the new form, the number of omission errors increased slightly, but the number of wrong entries dropped. Table 3 breaks down the errors into the three error types for both versions of the form and both levels of experience. Recall that omission errors involve leaving a required box blank, misplaced entries involve marking an irrelevant box, and wrong entries involve marking a relevant box incorrectly. A $3 \times 2 \times 2$ analysis of variance revealed a significant main effect of error type, $F(2,16) = 75.61$, $p < .01$. As Table 3 indicates, omission errors were the most frequent overall for both versions of the form, followed by wrong entries and then misplaced entries. Figure 3 shows the changes in the distribution of errors separately for the novices and the experienced subjects.

The most important difference in the distribution of errors was the sharp decrease in the number of wrong entries on the new version of the form. As the TYPE 3 rows of Table 3 indicate, when novices used the old form, they made an average of 8.33 wrong entries, but when they used the new form, they only made three errors of this type. Similarly, the number of wrong entries committed by experienced subjects dropped from 9.66 on the old form to 3.67 on the new form. A contrast was performed on the number of wrong entries by

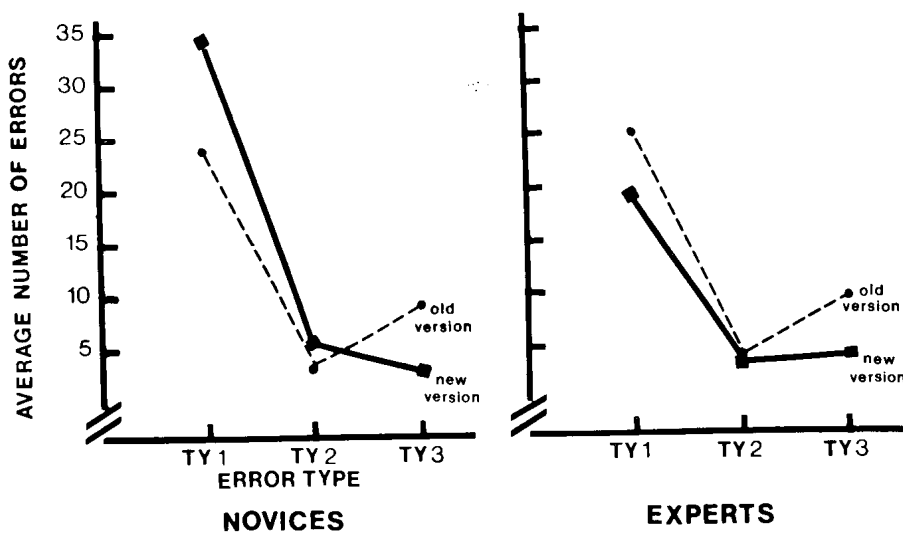


Figure 3. Average number of errors per task, by type of error, for novices ($n = 6$) and experts ($n = 6$). Type 1 = omission errors; type 2 = misplaced entries; type 3 = wrong entries.

version, collapsing over the experience variable. It revealed that subjects overall made significantly fewer wrong entries on the new form than on the old, $t(8)=2.165$, $p < .05$. As the later discussion of the thinking-aloud protocols will suggest, the increased use of checkboxes in the new version is a likely cause of the decrease, since checkboxes make encoding a response easier.

Looking at just the omission errors, it appears that novices left more relevant boxes blank on the new form than on the old, while experienced subjects did the reverse, committing fewer of these errors on the new version. These two trends apparently balance each other out. A contrast performed on the number of omission errors by version, collapsing over the experience variable, revealed no effect of version, $t(8) < 1$.

To sum up the results so far, subjects who used the new version of the form tended to complete it more quickly than subjects using the old version. While using the new version did not lead to fewer errors overall, the nature of the errors was less serious, since the number of wrong entries decreased for both novices and experienced subjects. The fact that novices made more omission errors and misplaced entries on the new form than on the old suggests that the new form does not solve all the novices' problems with the form, but alters how these problems are manifested. A novice who does not understand how to fill in a box on the new form is more likely to leave it blank, committing an omission error. Conversely, novices were more likely to fill in an entry on the old form, entries that were often incorrect.

Evidence from thinking-aloud protocols

The statistical results suggest that the new version of the form is easier to use than the old form, but because the differences between the two versions were not controlled systematically, they do not tell us why. The comments subjects uttered aloud as

they filled out the forms provide more detailed information about what aspects of the forms confused the subjects, or caused them difficulty. The protocols then helped us verify that we identified real problems in the old version of the form. The protocols also showed us that we found effective solutions to some of those problems in the new version of the form, and that we also created a few new, unexpected problems. Thus, we used the protocols to confirm suspected trouble spots on the forms, and identify new ones.

This section briefly reviews the problems that were identified in the old version of the form, and presents excerpts from the protocols that relate to them. The excerpts are illustrative of the problems subjects encountered, but they cannot be considered conclusive evidence. In order to draw conclusions from the protocol data, it would be necessary to code every comment in the protocols and do a quantified analysis of the types of comments, an analysis that goes beyond the scope of this paper.

Negative language

As we anticipated, many of the subjects had problems understanding the language on the old version of the form. For example, the two boxes in the Duplicating section (see the top portion of Figure 1), which imposed negative conditions ('Indicate only if not', 'If other than 20# Bond White used'), were difficult for the subjects to parse, and the subjects made a number of errors in these boxes, leaving them blank when they should have filled them in, or writing in them when it wasn't necessary to. The protocols suggest that even when they treated the boxes correctly, the subjects found the negative language hard to comprehend. Here are some of the comments three subjects made as they worked on the 'Indicate only if not' box:⁴

● Nov2: 'Indicate only if not ... only I don't understand that ... Indicate only if not ... only if not 8½ by 11 original? Oh, I see.'

● Nov3: 'Indicate only if not ... 8½ by 11, 8½ by 11, Black . . . I have no idea what that means . . . Indicate only if it's not these sizes. Oh I see.'

● Exp5: 'Indicate only if not ... black. It's not black [looking at the set of originals]. Oh, the ink, the letters? Yeah. I'll leave that blank. If it's not 8½ . . . , indicate it. Ok.'

The purpose of the negative language was to establish standard 'default' options. Clients fill in these boxes only if they want something other than the standard size and quality of paper, and the standard colour of ink. On the new version of the form, these boxes were rephrased in positive terms. As a result, the parsing problem was eliminated at the cost of requiring clients to fill in the boxes on every job, even if the option desired is standard. In fact, two of the experienced subjects were surprised at having to fill in these boxes on the new form when they wanted the standard options. The cost of increased clarity does not seem very great, however, when we recall that subjects filled out the new version of the form more quickly than the old, even though they had this additional writing to do.

Technical terms

Technical terms that appeared on the old version of the form also caused problems for the subjects, who had difficulty finding the right option even though they often could describe what they wanted in ordinary language. This problem was especially apparent in the Bindery section.

The protocols indicated that subjects had two problems when they tried to request a plastic spiral binding (a 'GBC'), for the Report Task. First, the types of bindings are listed under 'Stitching', but subjects did not associate the term 'Stitching' with a plastic binding. Subjects had the same problem in another task when they wanted a report stapled. They seemed to assume that 'Stitching' required needle and thread. The second problem that subjects faced was that none of the boxes, either

under Stitching or elsewhere, clearly referred to a plastic binding or to staples.

Here are the comments of two subjects as they worked on the Bindery section for the Report task. The first subject, Exp4, ignores the boxes under Stitching altogether because she assumes they have nothing to do with the kind of binding she wants. She tries to find the right option by guessing at another unfamiliar term, and then concludes that none of the boxes in the Bindery section is appropriate.

● Exp4: Do I have to do—OK, Bindery, AIS. [Points to the plastic binding.] Is there information in here about what this kind of bindery is called?

[The experimenter replies that she can only define terms that appear on the form.]

OK. Now AIS is additional information . . . Additional Instructions Sheet. I do want them to collate the items. [Checks the 'Collate items' box.] I'm not sure what I would write there exactly—'collate in order'? Stitching, Folding, Other. What's an NCR pad?

[Experimenter tells the subject that an NCR pad makes copies without carbon paper.]

OK, well that's not what I want, and I don't want it to be stitched. I guess then that there's an AIS required . . . But I still have to know what this is. I'll just say 'plastic rings'.

[Subject writes, 'Binding: plastic rings' on the Additional Instructions Sheet.] There probably is a name for what that is called. But—and it's probably not plastic rings, but again I would be talking to these people and I would tell them to go like that one [indicating plastic binding].

The following excerpt is from a subject who finds the right box by the process of elimination. She asks for definitions of terms in the Stitching section until she hits on the right one. Even after she has the right term however she has some trouble deciding how to indicate her choice. At first, she thinks that the neighbouring boxes 'Upper left' and 'Left

Margin', describe where the binding will go. Like some of the other subjects, she does not immediately associate 'Upper Left' and 'Left Margin' with staples. Her troubles here reflect both language problems and grouping problems that we attributed to the grid format, which will be discussed below.

● Exp5: 'Bindery—ooh. Should I attach AIS [Additional Instruction Sheet]? I don't know what this squiggly thing [indicates plastic spiral binding] is called.

[Subject asks successively for definitions of 'Saddle Stitch', 'Velo' and 'GBC'. Experimenter defines each of these terms, finally telling her that a GBC is a plastic spiral binding.]

Exp5: Oh. What does GBC stand for?

DC: It's just a brand name. I don't know what it stands for.

Exp5: How is a normal person supposed to know that? God. Ok. Ah, that was lucky. Upper left. Upper left what? The whole thing—I want the whole thing done. Left Margin—yes. 1, 2 and 3. I don't know what they're talking about. Left margin. Well let me first mark Stitching. Ok. GBC—yes, I want that done. What do I put—'on all of them'? I'll just put an 'x'—GBC. I don't know what Upper Left is. Oh I know, that's probably a staple in the upper left, probably. Well, whatever. I'll just put GBC. That was easy enough.

The combination of the grid format and the technical language made the Bindery section hard to fill out. Separating the stapling options from the types of binding, and simplifying the vocabulary on the new version of the form seemed to help subjects isolate the category they wanted (collating, stapling, folding etc), but we did not solve all of the problems. In particular, the terms under these categories (such as 'saddle stitch'), were still confusing.

Ambiguous language

Two boxes in the Duplicating section of the old form caused subjects more trouble than anything

else on the form: 'Number of Pages One Side' and 'Number of Pages Two Sides'. Not only are the labels on these boxes ambiguous, but filling in the boxes correctly requires a fairly abstract calculation. The information that the CPO staff wants from the 'No. Pages 1 Side' box is the number of sheets in each finished copy that will have printing on just one side. Similarly, the 'No. Pages 2 Sides' is the number of sheets in the finished copy that will have printing on both sides. In effect, these boxes require the client to project how many sheets of paper will be used in each duplicated copy. Adding together the numbers from these two boxes should yield the number of sheets in each finished copy, a number that might be very different than the number of originals the client is submitting. From the CPO's point of view, the information in these boxes is very important. The stockroom clerk needs to know how many sheets of stock to pull and how many to charge to the client. The client, on the other hand, only wants a way to tell the CPO which sheets to leave blank on the back.

An examination of the errors on the completed forms revealed that every subject who worked on the old form filled in these boxes incorrectly for at least two out of three tasks. (The fourth task did not require subjects to work in the Duplicating section at all.) The protocols indicate that subjects were confused by the labels on these boxes and rarely interpreted them as the CPO intended. In fact, the subjects came up with a number of different interpretations:

● Nov2: 'Oh, here's where No. Pages One Side becomes pertinent. Hum. I don't know to fill this out . . . all these pages [indicates originals] are one-sided but I want some of them to come out to be two-sided. Number of pages one side . . . that I want or that I have?'

● Exp4: 'Now I've got to figure out a way to . . . OK, title page . . . [Subject writes "Title page" in Items box.] Title page—well, number of originals—one. I want 250 of them. Number of pages one side—I guess that's 250

also. Yeah, because that's how many it would turn out to be.'

- Exp5: 'I want ten originals. I want 250 of each ten. Yes. Number of pages on one side. One, two,—oh, I think they all are. Number of pages on one side. Looks like every one of them. Ok. Ten.'

- Nov1: 'It's two pages but it's four sides, so we'll say it's four pages.'

- Exp4: 'Number of pages one side. There are no pages—it's an envelope!'

These two boxes were replaced on the new version of the form with a simple choice between single-sided or duplex printing. This change probably accounts for a substantial part of the decrease in the number of wrong entries committed on the new form. Subjects always tried to fill in these boxes on the old form and usually what they wrote was incorrect. However, on the new form, subjects generally chose the right option, or, if they were unsure, left both options blank. At worst then, subjects on the new form committed omission errors on these boxes, while subjects on the old form committed the more serious type 3 errors. The consequence of the change for the CPO is that a staff employee must now do the calculation instead of the client. This seems to be a reasonable arrangement. Assuming that the CPO clients made as many errors on these boxes as were committed under the experimental conditions, it is probable that the staff was already recalculating these boxes anyway.

Grid format

There is also evidence that the grid format of the old form caused subjects problems. The shading of the grey areas was intended to signal the subjects not to write in those areas. However, half of the subjects who worked on the old form ($n=3$), attempted to fill in the large grey area beside the

stock lines anyway. There seemed to be a strong tendency among the subjects to work all the way along a row from left to right, a tendency which was also noted in Wright and Barnard's (1978) subjects. So when the subjects finished writing in the stock descriptions, they just kept going—right into the grey boxes. Eventually, they realised that the grey area was not for them to fill out.

The subjects who worked on the old form found that the grid provided too little space for writing the necessary information. The following two comments are typical:

- Nov1: 'This box is too little to write the colour of ink in.'

- Exp4: 'There's not a lot of space to write anything in, especially here where they want to know what the stock is. There's all this stuff over here that I don't care about (indicates grey area) and I'm assuming that I don't ever have to write in there, but where I do want to write stuff I don't have any space.'

The layout of the new form seems to have solved the space problem. While four of the six subjects who filled out the old form complained that they had too little room to write in, not a single subject working on the new form voiced a similar complaint.

The protocol evidence, in combination with analyses of the errors subjects committed, suggests that the original assessment of the old version of the form successfully identified its major trouble spots, and that many of these were eliminated in the redesigned version of the form. The new version of the form did not solve all of the subjects' problems; subjects still commit errors when they fill it out on their own, but these errors are arguably less serious than those that usually appeared on the old version of the form.

Conclusion

This paper reported on a project to evaluate and redesign a work order and to test whether the

changes improved the clients' ability to complete the form correctly. Two methods were used to test the two versions of the form.

An experimental methodology was used to test whether clients would fill out the redesigned form more quickly and accurately than the old version. Since it is not difficult to measure how long subjects work on a task, nor to score a completed form for correctness, this method proved to be a relatively quick and painless way to obtain reliable evidence that the new version of the form would be easier for most people to fill out on their own. In addition, the quantitative data that the experimental method provided was useful for identifying items on the forms that were consistently filled out incorrectly.

The experimental method was combined with a more time-consuming, qualitative method: the analysis of verbal data from thinking-aloud protocols. Protocols usually cannot be used in a large-scale test because they rapidly produce too much data to deal with. However, because of the small number of subjects in this study, it was feasible to use protocols to study the process by which clients filled out the forms, and to discover what aspects of the forms caused subjects problems. By means of the protocols, we were able to identify (or confirm) problems that we would have missed by simply counting errors. For example, most of the people who filled out the original version of the form completed the 'Indicate only if not. . .' items correctly. However, the protocols revealed that these items did cause people trouble; they had to reread the labels on these items again and again in order to figure out what they meant. The protocols also helped us develop strategies for dealing with the individual problems on the form.

In sum, because of the pairing of these two methods, it was possible to use a very small sample population to demonstrate that the new version of the form was more useable than the old, and to obtain support for the particular changes we had

made. Combining methods in this way can help designers to reduce the high costs of frequent testing, in terms of both time and money, while at the same time obtaining more information than either method provides alone.

Notes

1 The author gratefully acknowledges the assistance of Carnegie-Mellon's Office of Publications and Campus Printing, especially Tim Basinski and Tom Kosak, the excellent advice of John R Hayes, and all those who read and commented on this manuscript.

2 The term 'box' will be used to refer to any place on the form where the form-filler indicates a choice or fills in information.

3 The rationale for thinking-aloud protocols derives from recent work in cognitive psychology and is discussed in Ericsson and Simon (1980), and Hayes and Flower (1983).

4 In the following excerpts, subjects will be referred to by their level of experience and by a subject number. So 'Nov2' is novice subject number 2, and so on.

Ericsson KA, Simon HA, 1980
Verbal reports as data
Psychology Review, 87/3, 215-251

Hayes JR, Flower L, 1983
Uncovering cognitive processes in writing: an introduction to protocol analysis
In Mosenthal, et al

Lefrere P, Macdonald-Ross M, Waller R, Abbott P, 1983
Effective forms: A case study of the development and testing of two postal claim forms for supplementary benefit
Institute of Educational Technology, The Open University, Milton Keynes

Matthews C, Jacobson M, Jones R, 1982
Common sense, skill and research in forms design
Information Design Journal, 3/2, 87-95

Mosenthal P, Tamor L, Walmsley S, 1983
Research in writing: Principles and methods
Longman, New York

Wright P, Barnard P, 1975
'Just fill in this form'—a review for designers
Applied Ergonomics, 6/5, 213-220

Wright P, Barnard P, 1978
Asking multiple questions about several items: the design of matrix structures on application forms
Applied Ergonomics, 7-14