Human Factors and Behavioral Science:

The *UNIX*™ Writer's Workbench Software: Results of a Field Study

By P. S. GINGRICH*

(Manuscript received December 17, 1981)

This paper describes a study in which the *UNIX*™ Writer's Workbench software was used by writers to analyze and revise the texts they wrote. For ten weeks, two groups of writers were observed: those for whom writing was their principal activity, and others for whom writing was only one of their responsibilities. These participants used the Writer's Workbench programs to produce documents related to their jobs. The results indicate that both types of writers can use the programs without changing their accustomed modes of operating. Writers found the programs helpful; they liked the immediate feedback the programs provided and the detailed suggestions on how to revise their texts. In addition, when editing prepared texts, writers found more errors using Writer's Workbench output than when they had no output available. Finally, participants thought they spent less time editing their documents when they used the Writer's Workbench programs, although total time spent on writing and editing did not change.

I. INTRODUCTION

The *UNIX*† Writer's Workbench software was developed in two major stages. The first version of the system contained 23 programs

---

* Bell Laboratories.
† Trademark of Bell Laboratories.

---

ranging from those that searched for proofreading errors to those that summarized stylistic features. These programs were used within Bell Laboratories for nearly a year, during which time data were collected on patterns of program use. A second, expanded version was then developed to reflect the experience gained from the first version.

To date, there is little formal evidence about user interaction with Writer's Workbench programs. We monitored the program on eight *UNIX* systems and observed that the programs were used most when they first became available, and then later they were used on a constant, but less frequent basis, amounting to an average of six to seven program runs per week per user. Informal discussions with users indicated that the peaks in use were due to users experimenting with a variety of input texts, good and poor, to evaluate the potential of the programs. Once they understood the potential, use dropped to a more normal level.

The present study examined the way experienced writers at two Bell System locations used the Writer's Workbench programs as a tool in performing their jobs. We were interested in examining how using the programs affected people's attitudes toward the system, the documents they write, and their modes of writing.

The Writer's Workbench programs provide feedback about many characteristics of writing. But do writers want the feedback? Which programs are used most often? Does using the programs make a difference in the way people write or in their attitudes about what constitutes good writing? How might the programs be modified to make them more useful to writers? This study was the first attempt to address these issues systematically.

II. FIELD STUDY

2.1 Background of the participants

Fourteen participants from each of two Bell System locations took part in the study. The 14 participants in Group 1 were course developers and instructors with a mean of 10.5 years service in the Bell System, and most had been in their current jobs between one and three years. Forty-three percent of the Group 1 participants had completed four years of college. Before the study began, these participants said they spent about 75 percent of their time writing and editing documents.

Of the 14 participants in Group 2, 11 were technical staff and 3 were clerks. For the technical participants, writing documents was only one of many job responsibilities; prior to the study, they reported spending about one-third of their time writing and editing documents. They also helped to design, develop, and assess systems. The three clerks were included because they processed documents for three managers

from whom we collected data. The technical participants had a mean of 13.6 years service in the Bell System, and most had been in their current jobs between one and three years. Sixty percent of these participants had completed four years of college, and half of those had master's degrees. All three managers had completed four years of college; none of the clerks had.

The two groups of writers differed in their approaches to writing documents and their knowledge of the *UNIX* operating system. Group 1 participants were more knowledgeable about the *UNIX* system and more inclined to compose, edit, and complete drafts on the terminal without the services of clerks. Such differences between groups could lead to different patterns of using the Writer's Workbench system.

2.2 Design of the study

2.2.1 Trial procedures

The study included two days of training, a one-week period for practice using the Writer's Workbench programs, and 10 weeks of using the programs in normal work activities.

During the two-day training sessions, participants completed questionnaires, learned about the Writer's Workbench system, and then used the programs. After we trained the participants, we tested them for mastery of 11 important concepts of the Writer's Workbench system, including characteristics of a document's style, specific program functions, and program options. Our training successfully conveyed the hierarchical arrangement of the major Workbench commands and the most frequently used options. Participants did have difficulty, however, in naming specific commands for given functions, perhaps because they had not yet used all the commands. All the commands, however, were explained during training.

After the training was completed, participants were encouraged to explore the Writer's Workbench programs during one week of warmup before the official on-line record keeping began. We provided structured exercises that participants could use to run each program, as well as a handbook that contained a tutorial introduction to the Writer's Workbench programs, and documentation describing each program.

For the next 10 weeks participants used the Writer's Workbench programs. We gave them guidelines, which suggested primary programs to run on each document they wrote, and we encouraged them to use additional programs as well.

2.2.2 Data collection instruments

2.2.2.1 Questionnaires. Participants completed questionnaires before the study began, at the midpoint, and after the study ended.

These questionnaires gathered demographic information about the participants, assessed their attitudes toward computers, and determined characteristics of their type of work and mode of writing. Some questions recurred on two or three of the questionnaires to measure changes in attitudes and modes of operating.

2.2.2.2 On-line data collection. Every time a Writer's Workbench program was run, the program automatically recorded its name, the date and time, who ran it, and the name of the text file it analyzed. Some programs, such as the prose analysis programs,[1,2] also saved data about the text characteristics. We used this information to describe program use across time and also to note changes in text files.

In addition, an interactive program prompted users to rate the helpfulness of each program's output on a five-point scale. These data were collected to determine users' opinions of how well the programs analyzed each text file.

2.2.2.3 Revision tasks. During the study, participants completed four document revision tasks, which involved editing in ten minutes a 300-word passage containing planted errors. Each participant edited the four different passages once. Each passage appeared equally often in the following sessions:

Session 1. Edit one passage before training begins (no Writer's Workbench program output).

Session 2. Edit one passage immediately after training (no Writer's Workbench program output).

Session 3. Edit one passage at the end of the study. (Appropriate Writer's Workbench program output was attached to each passage.)

Session 4. Edit one passage three weeks after the end of the study (no Writer's Workbench program output).

These revision tasks provided a controlled environment in which to compare the participants' ability to improve a text at various stages of the study.

2.2.2.4 Interviews with participants. Each participant was interviewed for 25 minutes at the end of the study. We asked 11 general questions to elicit the participants' comments on how they used the Writer's Workbench programs and what they liked most and least about the system.

2.3 Results

2.3.1 User acceptance of the programs

We examined user acceptance of the Writer's Workbench programs in several ways. Program use was recorded automatically, and following each use, participants rated the helpfulness of the output. In interviews, participants also reported what they liked least and best

about the system. We found no significant difference in the way the two groups of participants used the programs or their acceptance of them. Consequently, their data were combined in most of the following sections.

2.3.1.1 Program use. Although the way participants used the programs varied greatly, on the average they ran about six programs a week.

Table I shows the number of times each program was run during each week of the study. The wwb program, along with its two components proofr and prose, and the spellwwb program were run most often. Use of the programs diminished over the course of the study, but use of the informational "help" programs, such as punctrules and wwbinfo, dropped to zero some weeks before the end of the study. It seems that participants used the "help" programs only as they learned about the Writer's Workbench programs.

2.3.1.2 Helpfulness ratings. Participants rated the helpfulness of the output each time they ran a Writer's Workbench program and generally found the programs helpful. The average rating was 3.8 on a scale of 1 (not helpful) to 5 (very helpful). The wwb and prose programs, those most frequently used by Group 1 and Group 2, respectively, had average ratings of 4.0 and 3.7 on the helpfulness scale. Table II shows the mean rating for each program by group.

2.3.1.3 What participants liked best. During the final interviews and on questionnaires, many participants said that the Writer's Workbench programs were most valuable because they gave immediate, objective criticisms on concrete problems in their text. They liked the advice from the proofr component because they found it accurate and complete.

For the prose component, participants reported that they felt the advice was objective and specific, but they often were not sure of how to make the changes it suggested. Nevertheless, many participants reported that the stylistic information from prose was an improvement over the often vague advice and subjective opinions of human reviewers.

Participants commented on three other general aspects of the Writer's Workbench system. First, it saved time (both by the speed with which it reviewed text and by eliminating delays caused by looking for a colleague to review the document). Second, it was completely private, giving the writer a chance to improve the document before anyone else saw it. Third, participants reported they were more aware of principles of good writing and of how to change their style to suit the purpose and audience of a particular document.

2.3.1.4 What participants liked least. Two criticisms of the programs surfaced during final interviews and from questionnaire responses.

Table I—Average individual program use during training, warm-up week, and trial weeks

Program	TR	WU	Week of Trial											Total
			1	2	3	4	5	6	7	8	9	10	11†	
wwb programs														
*wwb	128	111	52	43	29	22	22	38	22	21	6	9	9	512
proofr	11	29	14	15	3	6	7	8	3	1	0	2	0	99
spellwwb	22	58	11	27	13	10	20	42	20	10	20	5	4	262
punct	17	10	2	0	0	1	0	1	2	0	0	0	0	33
double	9	4	0	3	1	1	0	0	1	0	0	1	0	19
dictplus	0	3	2	6	2	0	2	0	1	0	0	3	0	17
diction	0	4	2	0	0	1	0	0	4	0	0	0	1	15
suggest	1	23	21	0	1	1	7	10	0	0	2	0	4	68
splitinf	0	2	0	0	0	0	0	0	0	0	0	0	0	2
prose	44	47	22	25	9	8	31	7	9	13	2	4	0	221
parts	1	4	1	1	1	0	2	0	1	1	2	0	0	12
style	32	27	13	5	9	5	23	5	5	3	3	2	6	138
Help programs														
spelltell	9	3	2	8	2	1	1	6	8	0	0	0	0	40
worduse	14	22	17	6	5	11	16	8	2	1	0	0	5	107
punctrules	6	5	0	0	1	0	0	0	0	0	0	0	0	12
splitrules	6	3	1	0	1	0	0	1	0	0	0	0	0	12
wwbhelp	16	8	2	1	1	0	1	0	0	0	0	0	0	29
wwbinfo	26	4	1	2	1	1	2	0	0	0	0	0	1	38
wwbstand	10	4	0	1	1	1	1	2	1	0	0	0	0	21
Other programs														
*abst	10	12	3	53	12	0	0	13	1	2	0	0	0	106
*acro	5	7	2	0	1	3	0	0	0	2	0	0	1	20
chunk	0	4	1	1	0	2	1	0	0	1	0	0	0	11
dictadd	9	17	7	2	0	1	0	2	4	0	0	3	0	45
*findbe	16	26	12	12	4	0	3	2	3	2	1	1	2	83
match	4	4	0	2	3	0	3	2	2	1	1	3	1	24
org	2	3	0	1	1	0	5	0	0	1	0	0	2	15
*sexist	10	45	6	22	2	1	0	1	0	1	0	0	0	88
spelladd	20	13	6	10	2	28	2	7	1	0	0	0	0	87
syl	4	1	1	3	0	0	1	0	1	0	0	0	0	10
topic	0	2	1	1	0	0	1	0	0	1	0	0	0	6
wwbmail	6	3	5	11	2	1	1	2	4	1	0	0	0	36
Total	438	508	207	261	105	105	151	155	95	62	32	33	36	2188

* Programs whose use was highly recommended in the trial guidelines.
† Week 11 includes only use by Group 1.

Table II—Users' ratings of Writer's Workbench programs

Program	Group 1		Group 2	
	Mean Rating of All Participants	Number of Program Runs	Mean Rating of All Participants	Number of Program Runs
wwbhelp	NR*	0	5.0	3
punct	3.0	4	5.0	2
splitrules	3.5	2	5.0	1
punctrules	NR	0	5.0	1
spelladd	3.6	41	4.9	12
wwbinfo	NR	0	4.7	6
sexist	3.0	7	4.7	33
spelltell	2.0	3	4.7	23
wwbmail	NR	0	4.6	28
spellwwb	3.9	123	4.4	56
dictplus	NR	0	4.4	14
double	NR	0	4.2	6
worduse	3.3	43	4.1	27
wwb	4.0	156	3.9	92
proofr	4.0	3	3.9	41
prose	2.5	12	3.8	94
abst	2.1	16	3.7	67
style	4.0	42	3.7	23
wwbstand	NR	0	3.6	7
acro	2.7	3	3.6	5
findbe	3.1	21	3.4	8
dictadd	3.4	9	3.4	8
match	2.5	8	3.0	4
topic	4.0	1	3.0	2
org	2.9	8	3.0	2
parts	2.3	4	2.7	3
chunk	1.0	3	2.7	3
syl	NR	0	2.3	4
diction	2.4	8	2.0	34
suggest	2.9	42	NR*	0
splitinf	NR	0	NR	0
Average (Total)	3.6	559	4.0	578

* NR means no rating.
NOTE: The number of ratings does not agree with the number of programs run shown in Table V because participants did not begin rating program output until after the warm-up week. Also, some end data was not incorporated from the last week.

First, some participants reported that they found it difficult to apply the advice they received from the prose and diction programs. They either were unfamiliar with the principle the program checked or were unsure of how the principle was being evaluated.

Participants also criticized the length of some program outputs. For example, some said that for the wwb command, the default output was too long and wwb − s output was too short. An output of intermediate length was suggested.

2.3.2 Effects on documents

What effects does using the Writer's Workbench programs have on the documents produced? Do writers locate more errors using Writer's Workbench output to edit a document? Does using the programs for

a period of time later enable writers to edit more effectively on their own?

2.3.2.1 Revision tasks. Results from the revision tasks show how the participants' ability to improve a document differed at various stages of the study. The revision tasks were completed: (1) before the study, (2) after training, (3) after 10 weeks of using the Writer's Workbench programs, and (4) three weeks after the study ended. Output from the Writer's Workbench programs was available only in the third session.

Passages used in the revision tasks were adapted from stories in the *Bell Laboratories Record* and traced the development of new Bell System operations. The average length of the passages was 300 words. Table III describes the twenty errors planted in each of the four passages.

Of the nine categories of inserted errors shown in Table III, the Writer's Workbench programs cannot provide feedback on using a word that looks like the correct word (category g), mistyping one word that produces another (category h), and making a mistake in the agreement of subject and verb (category i).

Table IV shows the results of the revision tasks. More errors were detected in Session 3 than in any other session. Having the Workbench output available improved performance compared both with where the user had no experience with it (Sessions 1 and 2) and where the user

Table III—Type and number of errors in each revision passage

Type	Number of Error(s)	Item
a	4	Spelling errors, which were not legitimate words and would be found by the spell program
b	2	Punctuation errors that the punct program could identify
c	6	Wordy phrases included in the diction phrase dictionary
d	1	Immediate repetition of a word
e	1	Split infinitive
f	3	Instances of sexist language
g	1	Case of a word substituted for another that looked similar
h	1	Misspelled word that was a proper English word and would not be found by the spell program
i	1	Case of subject verb disagreement

Table IV—Proportion of errors detected for each session*

	Session 1	Session 2	Session 3	Session 4
Mean	0.271	0.275	0.408	0.309
Standard error	0.006	0.007	0.007	0.007
Number	21	21	21	21

Session 3 vs Session 1—$t_{(20)} = 3.35, p < 0.005$
Session 3 vs Session 2—$t_{(20)} = 3.41, p < 0.005$
Session 3 vs Session 4—$t_{(20)} = 2.49, p < 0.05$

* Data are reported for those 21 participants who completed all four sessions.

had employed the Writer's Workbench system in the past but did not have its output available while proofreading (Session 4).

The data are broken down by category in Fig. 1. We see from this figure, which shows the percentage of detected errors for each session (bars 1 to 4) for each category, that Session 3 is clearly distinct from the others.
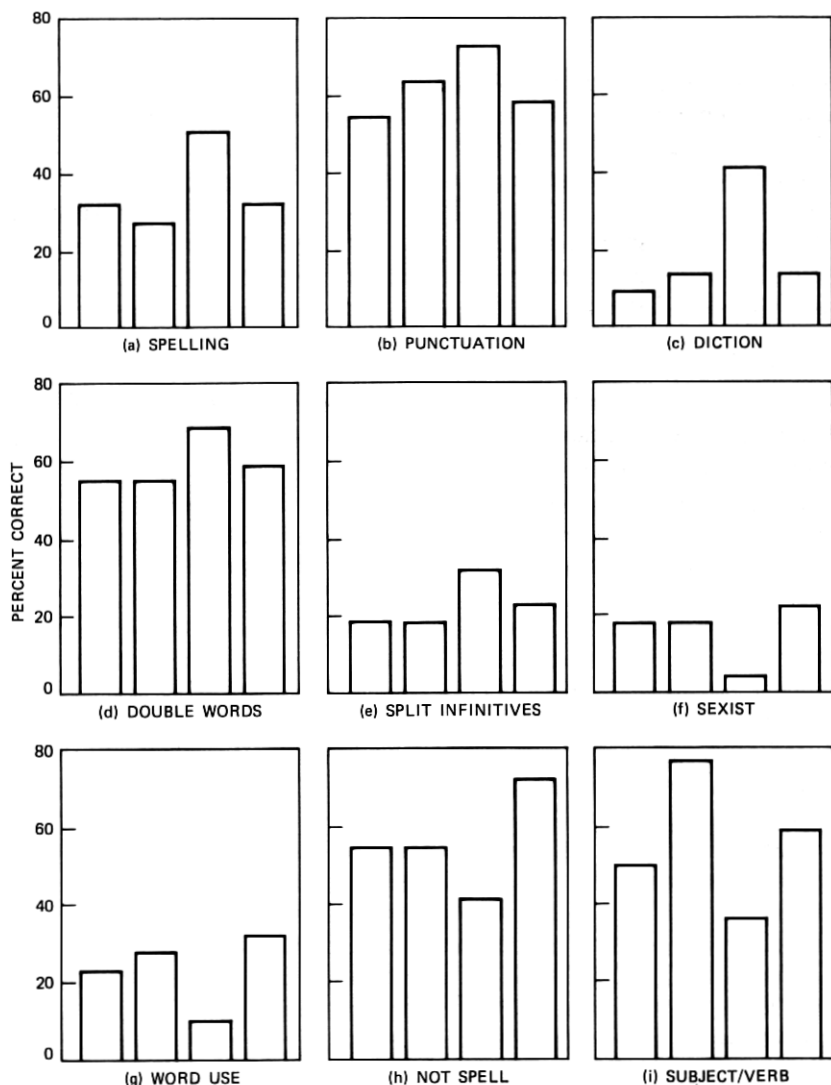

Fig. 1—Percentage of errors detected for each error category (a through i) in Sessions 1 through 4: (a) spelling errors; (b) punctuation errors; (c) wordy diction; (d) double words; (e) split infinitives; (f) sexist language; (g) incorrect word use; (h) spelling errors not found by spell program; and (i) subject/verb disagreement.

We included the errors in categories g, h, and i because we wanted to determine whether the participants had become so dependent on the Writer's Workbench analysis that they would fail in Session 3 to notice the errors that it does not find. The results of Session 3 support this supposition; participants found many of the Writer's Workbench errors and overlooked others.

It appears that participants used the time in Session 3 to find and mark the errors given by the Writer's Workbench programs, but did not have a chance to proofread the paper for other errors. Indeed, several participants remarked that they spent their time reviewing the output and locating the corresponding error in the passage, rather than proofreading the passage as a whole. As a result they may have run out of time. This seems likely in view of the low rate of marking sexist usage, the output for which came last in the collection of Writer's Workbench program outputs given to participants. As Fig. 1f shows, participants marked far fewer sexist usages in Session 3 than in any other session. Even so, overall performance was best in Session 3.

2.3.3 Effects on writers' activities

The writers completed questionnaires on how the Writer's Workbench programs influenced the way they prepared documents and assessed their stylistic features.

2.3.3.1 Mode of writing. The questionnaires asked participants how they write: Do they write on-line or on paper? Do they make their changes on-line or on paper? As we described above, before the study began participants in Groups 1 and 2 differed in their methods of writing and editing. To test whether their methods changed during the study, we performed a 2×3 (group $\times$ time) analysis of variance on each of these two items. (The range of responses to both questions represented the degree of involvement with a terminal during the task of either writing or editing. A value of 1 represented the most direct involvement with the terminal and a value of 4 represented the least.) Again, time (pre-, mid-, and post-study) was the within-subject repeated measure; and group (Group 1 or Group 2) was the grouping factor. The results show only a significant group effect for the writing method $[F(1,15) = 7.32, p < 0.05]$, as well as for the editing method $[F(1,15) = 21.0, p < 0.001]$. Participants in Group 1 were less likely to have a clerk do the terminal work for them ($M = 1.43$ for writing and $M = 1.19$ for editing) than were participants in Group 2 ($M = 2.47$ for writing and $M = 2.37$ for editing). There were no significant time effects or time $\times$ group interactions.

2.3.3.2 Estimates of stylistic features. We wanted to know whether using the Writer's Workbench programs would affect the way participants evaluated the stylistic errors in their documents. The pre-,

mid-, and post-questionnaires included questions about nine features of writing style that are measured by the Writer's Workbench programs. Participants estimated whether their draft documents typically had

—too few

—the right amount

—too many instances

of each feature, or they indicated that they could not evaluate the amount by saying either

—there was no right or wrong answer, or

—they could not answer the question.

The most marked difference was in how participants evaluated their proportion of passive sentences before and after the study. Before the study, about 85 percent of the Group 1 participants said there was no right or wrong number of passives sentences, or said they could not answer. After the study, no one answered in either of those categories; half thought they had too many passive sentences in their drafts, and half thought they had the right number. Before the study, more than 50 percent of Group 2 said they either had the right number of passive sentences, or they could not answer. By the end of the study, 60 percent of Group 2 thought they either had too many, or too few, passive sentences.

2.4 Discussion

Participants were able to use the programs frequently, liked them, and found the output helpful. By reviewing the results of the different measures together, we come to a clearer understanding of what effects the Writer's Workbench programs have on writers.

First, the data on program use reveal that, on the average, participants ran six Writer's Workbench programs per week. (This figure is comparable to data collected from Bell Laboratories *UNIX* systems we have monitored in the past.) Over time, there was a reduction in the total number of programs run, possibly because the novelty of using the programs wore off, or because of the cyclical nature of the documentation process. Writing a document generally entails collecting information, organizing it, and then writing, editing, and revising. Participants may have confined their use of the Writer's Workbench programs to the first stage of the editing process and to the final version. In addition, documentation efforts vary with the phases of product development; thus, the activities that the Writer's Workbench programs can support are cyclical. Observing program use over a much longer time period, perhaps a year, would give a clearer understanding of the long-term frequency of use, one less subject to seasonal cycles of writing demands.

Even when program use declined over time, the programs were still rated as helpful. The average helpfulness rating was 3.8 on a scale from 1 (not helpful) to 5 (very helpful). The use of programs was more strongly determined by individual needs than by our recommendations. For example, use of the acro program, which locates and prints all acronyms in a document, was highly recommended in the study guidelines. Yet, users in Group 1 did not find this program helpful (2.7 helpfulness rating) because it gave them too much output. Hence, no one used it after the third week of the study. The participants did, however, continue to use the programs that took the tedium out of editing, such as the proofreading programs. That the participants continued to use many of the programs indicates that they found these programs helpful.

The results of the revision tasks clearly show that the programs are helpful in locating errors when the person is under pressures of time. Using the Writer's Workbench system in Session 3, participants found significantly more errors in the same amount of time than in the other sessions. To produce their typical error-free final products, participants would need to spend far more editing time without Writer's Workbench system than with it. The results of the Session 4 revision tasks also show that one must use Writer's Workbench programs continually to get this benefit; it is not enough to have used them in the past. The computer is just much better at finding certain errors than our participants were. However, we do not know whether error types not located by the programs would tend to be missed by the human proofreader more or less often as a result.

When we look at how using Writer's Workbench programs affects a writer's mode of writing, we see few changes. Those who used clerks to type their texts before the study continued to do so during the study. Those who typed their texts themselves continued to do so. Participants did not need to change how they wrote to use the Writer's Workbench programs effectively, nor did it increase or decrease the personal contact individuals had with terminals.

In the current study, we were limited in the degree of control we could impose on the participants. There were no rewards for using the Writer's Workbench programs and incorporating the suggestions into later drafts, nor were there any punishments for not using them. We could not control how much or how little participants wrote during the 10-week trial. We had no independent judges to evaluate the quality of the documents produced with the aid of the Writer's Workbench programs, nor did we have comparable people writing comparable texts without the aid of the programs so we could compare our participants' texts. Thus, we were limited in the issues we could address.

Currently, studies with university English composition students are addressing some of the unanswered questions. In composition classes, students are using the Writer's Workbench program output to edit and revise their essays. Instructors and independent judges will evaluate the essays and compare the quality of these essays to those of control students not using the Writer's Workbench programs. In addition, these new studies include unskilled writers rather than experienced adult writers, which will allow us to determine whether using the system helps students learn how to write better. These findings will enable us to evaluate more fully the effectiveness of the Writer's Workbench programs.

III. ACKNOWLEDGMENTS

REFERENCES

1. N. H. Macdonald, "The UNIX™ Writer's Workbench Software: Rationale and Design," B.S.T.J., this issue.
2. N. H. Macdonald, L. T. Frase, P. S. Gingrich, and S. A. Keenan, "The Writer's Workbench: Computer Aids for Text Analysis," IEEE Trans. Commun. (Special issue on communication in the automated office), *COM-30*, No. 1 (January 1982), pp. 105–10.

AUTHOR

Patricia S. Gingrich, B. S. (Education), 1970, Pennsylvania State University; M.S. (Education), 1973, Hunter College of the City University of New York; Bell Laboratories, 1979—. Before joining Bell Laboratories, Ms. Gingrich taught at Hunter College. At Bell Laboratories she worked on the development and testing of the Writer's Workbench programs. She is currently a member of the Product Strategy Department, where her work includes product evaluation and planning.