HUMAN FACTORS IN COMPUTER SYSTEMS: GAITHERSBURG CONFERENCE CONTINUED
(Review of Session on Evaluating Text Editors)

W.J. Hansen
Three Rivers Computer Corporation

Evaluating Text Editors (Session 3B), chaired by Henry Ledgard, Human Factors Limited, and John Whitside, Digital Equipment Corporation. Text editing is probably the one activity that users spend the most time performing, so it is appropriate to devote considerable attention to improving its human factors. This session covered several techniques for evaluation of text editors.

"Evaluation of Text Editors" by Teresa Roberts, Xerox Systems Development Department, and Thomas Moran, Xerox Palo Alto Research Center.

The evaluation methods reported in this paper can be carried out by non-psychologists with a minimum of equipment and preparation. Four factors are measured for each editor: (a) time to perform a specified set of edits (b) frequency of errors by experts with the editor (c) learning time for novices, and (d) functionality, the possibility of performing a specified set of tasks easily within the editor's command set. (The set of tasks for the functionality measure is itself a valuable contribution to the design of text editors.) The results reported for eight text editors accord well with my own intuitive evaluations of these editors, so the methodology does seem to be producing valid results. In general, the full screen editors evaluated at twice the speed of the line-oriented editors: the most complex editor, TECO, was slowest to use, even for experts at its use. The fastest editor, Gypsy, uses only a mouse and a five key handset, so it scored the lowest score for functionality (though only slightly lower than TECO.) When I plotted time versus functionality for the full screen editors, I found that the fuller functioned editors were indeed slower, though only by a small margin. Because it provides a methodology that can be widely and usefully applied, this paper--and the thesis on which it is based--merits a wide readership.

"An Ease of Use Evaluation of an Integrated Document Processing System" by Michael Good, Massachusetts Institute of Technology.

At the time of the experiment reported in this paper, the Etude system was merely a prototype, so there were no expert users available to apply Robert's methodology. Instead, subjects were trained on Etude and performed a short task on both Etude and a typewriter. In the event, the typewriter proved faster for the task. However, the paper is able to show that the users, none of whom has prior computer exposure, could use the Etude system without anxiety and with positive attitudes.

"An Analysis of Line Numbering Strategies in Text Editors" by M.L. Schneider, S. Nudelman, and K. Hirsh-Pasek, Sperry Univac.

In "fractional line numbering," 1.3 is the same as 1.30 and follows 1.25; in "hierarchical line numbering," 1.3 precedes 1.25 and intermediate numbers include additional levels of hierarchy: 1.3.1, 1.3.2,....

Fractional line numbers were shown to be faster to use, even though the subjects were familiar with hierarchical numbers. (Although this result may be of limited application to screen editors, it may imply that section identifiers for hierarchical text ought not to be purely numeric.)

"Can We Expect to Improve Text Editing Performance?" by David Embley and George Nagy, The University of Nebraska-Lincoln.

A better title for this paper might be 'How we are trying to evaluate text editor performance.' The paper reports the current status of a developing laboratory for analysis of editor usage. This laboratory includes instrumenting editors to capture command-by-command times. A real innovation is the capturing of data from students as they are using the editor for course work. Such field experiments can go a long way toward validating the results of more limited experiments. Although not in the paper, the presentation at the conference reported on a technique for evaluating the cognitive load of individual commands. An analysis of the matrix of transition times from one command type to another can serve to divide the inter-command time into time for absorption of the result of one command and time to prepare for the next. It will be interesting to see results from this analysis when the work is complete.