Comparing Response Time, Errors, and Satisfaction Between Text-based and Graphical User Interfaces During Nursing Order Tasks

2 authors, including:

Nancy Staggers
University of Utah
42 PUBLICATIONS **783** CITATIONS

SEE PROFILE

Original **Investigations**

JAMIA

Research Paper ■

Comparing Response Time, Errors, and Satisfaction Between Text-based and Graphical User Interfaces During Nursing Order Tasks

Nancy Staggers, RN, PhD, David Kobus, PhD

A b s t r a c t Despite the general adoption of graphical users interfaces (GUIs) in health care, few empirical data document the impact of this move on system users. This study compares two distinctly different user interfaces, a legacy text-based interface and a prototype graphical interface, for differences in nurses' response time (RT), errors, and satisfaction when the interfaces are used in the performance of computerized nursing order tasks. In a medical center on the East Coast of the United States, 98 randomly selected male and female nurses completed 40 tasks using each interface. Nurses completed four different types of order tasks (create, activate, modify, and discontinue). Using a repeated-measures and Latin square design, the study was counterbalanced for tasks, interface types, and blocks of trials. Overall, nurses had significantly faster response times ($P < 0.0001$) and fewer errors ($P < 0.0001$) using the prototype GUI than the text-based interface. The GUI was also rated significantly higher for satisfaction than the text system, and the GUI was faster to learn ($P < 0.0001$). Therefore, the results indicated that the use of a prototype GUI for nursing orders significantly enhances user performance and satisfaction. Consideration should be given to redesigning older user interfaces to create more modern ones by using human factors principles and input from user-centered focus groups. Future work should examine prospective nursing interfaces for highly complex interactions in computer-based patient records, detail the severity of errors made on line, and explore designs to optimize interactions in life-critical systems.

■ **JAMIA.** 2000;7:164–176.

Informaticians and designers think that a key barrier to user acceptance of computers is the lack of user friendliness.[1–3] As evidence of this notion, Donald Lindberg,[4] Director of the National Library of Medicine and of the National Coordination Office for High Performance Computing and Communications, identified one of his major challenges as achieving a better fit between computers and the way people work. Orthner[5] called the interaction between the provider and the computer interface an area in desperate need

of attention. As early as 1992, the National Institute for Nursing Research cited the "man–machine interface" as a research priority for nursing informatics.[6] The challenges issued by these authors are echoed throughout health care today: The need for research on human–computer interaction is imperative.

The costs of not accommodating a specific discipline's requirements in computing are high, and loss of productivity and increased errors with poorly designed user interfaces are typically reported.[7] More work is needed to determine which type of display interface would be optimal for user efficiency, effectiveness, and satisfaction in system use. This study compares the differences in nurses' response times (RTs), error rates, and satisfaction ratings when they used a text-based interface and a prototype GUI to perform nursing order tasks.

Literature Review

Health care has been slow to adopt usability techniques that have long been used by such corporations as Xerox, Microsoft, Bell Labs, and Apple. These techniques formally assess the "user friendliness" of a system, its efficiency and effectiveness, and the user satisfaction in completing job-related tasks. Several calls to action were issued to incorporate these techniques into the design of health care systems.[3,5,8,9] The first documented usability laboratory for health computing began at the Mayo Clinic in 1997[10]; however, few others in health care have followed suit.

Developers typically provide a general accounting of how screen design occurred in a project; an example is reported by Sittig et al.[11] Explanations of systematic methods used to determine user requirements and interface design are less common, despite Patel and Kushniruk's excellent review[12] of cognitive methods for design determination and other useful techniques.[13] A systematic approach to both design and evaluation, including usability testing, is needed to determine health care computer interface requirements.

The general trend in computing has been to shift to GUIs.[14,15] However, available empirical evidence of their effects on users' productivity is mixed. Tomaiuolo[16] evaluated end-user perceptions of a GUI in online literature searches and found that users had positive perceptions of GUIs. Several authors[17–19] found that subjects performed better when they used GUIs for database searches, spreadsheet applications, and word processing. Rauterberg[20] evaluated menu selection using a text-based interface and a GUI for both novices and experts. Users were faster with the GUI; experts needed 51 percent less time to complete tasks. Davis and Bostrom[21] examined the differences between a command-based (text) interface and a Macintosh GUI during the performance of file directory and structure utility tasks. They found a significant improvement in the ability of novice users to learn and perform with the GUI.

In a series of studies, Temple, Barker & Sloane, Inc.[22] compared the word-processing and spreadsheet task activities of 120 white-collar workers using a GUI and a text-based interface. Using a between-groups design, they found that users completed 35 percent more tasks, were 17 percent more accurate (91 percent vs. 74 percent), were less frustrated and less fatigued, and learned more capabilities with the GUI than with the text-based interface. The authors asserted that their research findings support the theory that a GUI facilitates navigation within an application and among applications. That is, instead of navigation by the stilted commands and menus used in text systems, GUIs allow direct navigation between applications by the use of icons and within applications by the use of buttons.

The results of the study suggested broad benefits for individuals and organizations when navigation across various processes is required. Difficulties with the study include nonequivalent software applications with a GUI that had more functions than the text-based interface (although the GUI was faster despite its more complex functions), the lack of data about controls for individual differences, and the expectation that novices would teach themselves the applications after a brief introduction to computers.

In contrast, Whiteside et al.[23] compared outcomes for users of three different systems—command, menu, and iconic (GUI) systems. They found no significant differences in performance between the three formats and a performance degradation when novices used the GUI interface. D'Ydewalle et al.[24] also discovered that experienced users of word-processing programs performed less well with a GUI than with a text-based application. Carroll and Mazur[25] found that experienced command-based users learning Apple applications had problems selecting icons, opening and closing system files, and creating documents.

Several issues are apparent across these studies. First, the results of the studies are seemingly contradictory but might be explained by differences in the types of tasks and the study methodologies. Some authors used moderately complex tasks.[16–19] For example, Davis and Bostrom[21] were interested in the learning aspects of tasks performed with a GUI, while Tomaiuolo[16] studied

both simple and more complex bibliographic retrieval tasks. The Temple group[22] studied navigation within an application, whereas Whiteside et al.[23] investigated simple filing tasks. Rauterberg,[20] in fact, found a significant interaction between novices and experts and type of task. These differing results may be highly dependent on the complexity of the tasks and indicate the importance of empirical testing before a different display style is incorporated into a system.

Findings might also be explained by uncontrolled individual differences in the between-groups studies, in which user characteristics such as gender, past computer experience and, in particular, amount of practice with the system may have influenced dependent variables. Authors of the studies mentioned here were careful to examine differences for novice and expert computer users, but the groups may not have been practiced in the task at hand, confounding the observed differences. In addition, the results have limited generalizability, since women are under-represented in the reported samples. Finally, many of these studies were not published in peer-reviewed journals, only in conference proceedings. More rigorously reviewed studies are needed.

The current study addressed several deficiencies of previous studies. Participants included women and men, they were familiar with and had practiced the tasks at hand, individual differences were controlled by study design, and the sampling technique provided a better representation of the population of military nurses. The study required a range of tasks, from simple navigation to moderately complex tasks. Most important, the tasks represented realistic clinical nursing activities.

Conceptual Framework, Study Design, and Hypotheses

The conceptual framework for this study was the Staggers and Parks nurse–computer interaction framework.[26,27] The framework conceptualizes the nurse–computer interaction as a task-based information exchange between nurses and computers (Figure 1). Computers have characteristics, such as the time it takes to retrieve critical information from an application, that influence associated nurse behaviors. Nurses initiate tasks, and information is exchanged. Interactions occur along a development trajectory as user and computer mature, representing both a user learning effect as well as changes in technology.

This framework provides overall guidance to human–computer interaction studies. Besides considering the

differences among computer interfaces, an investigator needs to consider each factor of the framework in planning these studies. For example, controlling or studying individual differences (labeled as nurse characteristics) is necessary. Likewise, controlling for computer characteristics would lead an investigator to ensure that differences in computer processors did not affect results, and the developmental trajectory suggests the need to consider how much users have practiced the task at hand.

The hypotheses of the study were as follows: Nurses' response times with a prototype GUI will be faster than with a text-based interface across all types of computerized nursing orders; nurses' errors and incorrectly filed orders with a prototype GUI will be lower than with a text-based interface across all types of computerized nursing orders; and after the performance of computerized nursing orders, nurses' ratings of interface satisfaction will be higher with a prototype GUI than with a text-based interface. A counterbalanced repeated-measures design was used.

Methods

Sample

Approval to conduct the study was obtained from the institutional review board and the nurse executive of a large military hospital on the East Coast. As part of a larger study, the sample was obtained by a stratified (by gender), random technique. Of 127 randomly selected nurses eligible to participate, 98 nurses voluntarily completed the study, a 77 percent participation rate. Patient care demands precluded most of the remaining nurses from participating; a few simply declined to participate in the study. The mean age of participants was 33.9 years. They averaged 7.5 years
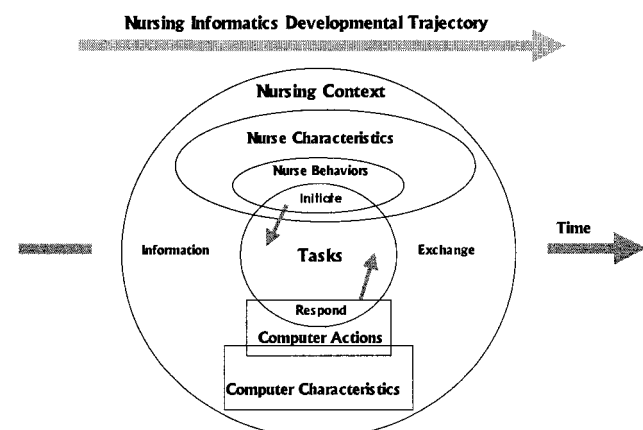

F i g u r e 1 Staggers and Parks nurse–computer interaction framework.[27] Used with permission.

in nursing and 8.7 years of military service. Fifty participants were men and 48 were women. The modal educational level was a BSN (69 nurses), but 28 had master's degrees and 1 had a doctorate. The sample included clinical staff nurses (77 percent), head nurses (2 percent) and nurses in other clinically related positions, such as instructors, advanced practice nurses, nurse anesthetists, case managers, and outcome managers (21 percent). The sample consisted of nurses in medical-surgical care (28.6 percent), intensive care (23.5 percent), operating room/postanesthesia care (12.2 percent), psychiatric care (11.2 percent), pediatrics (5.1 percent), and associated areas such as education and anesthesia nursing (14.3 percent). The reported unit and demographic characteristics were representative of the institution.

The Staggers Nursing Computer Experience Questionnaire, or SNCEQ,[28] was given to nurses to provide a description of their previous computer experience. In earlier work, a panel of nursing informatics experts assessed the content validity of the questionnaire; content validity indexes for items ranged from 0.83 to 1.00. Construct validity was estimated by comparing the computer experience of a group of informatics nurses to a group of graduate nurses. For the subscale and overall scores, ratings were significantly higher for informatics nurses. Test–retest reliability was completed with a group of 24 graduate nurses. Pearson product-moment correlations ranged from 0.90 to 0.97.

The mean computer experience for the sample as measured by the SNCEQ was 105 out of a possible score of 292, representing a low-moderate amount of experience. The majority of scores clustered in common personal computing applications (word processing, e-mail) and basic hospital information functions. However, the range of scores was wide, from 12 to 235. Operating room nurses and psychiatric nurses had lower mean scores than other groups. The sample as a whole used the installed text system primarily for e-mail and results retrieval; few ever used it for nursing orders, the task studied here.

Procedure

Nurses were tested in small groups in a quiet room away from patient care areas. They took approximately three to four hours to complete the entire experiment. All participants used identical computers, with 333-MHz Pentium II processors and 17-inch super VGA screens. The computers were disconnected from the local area network to prevent network interference with processing.

After reading and signing the informed consent, nurses completed the demographic information and the computer experience measures. They then received standardized instruction about the computerized portion of the study. Standardized training on both systems consisted of critical commands for each required action, e.g., to discontinue an order. Each nurse had a reference card showing the requested actions and commands, ensuring minimal memorization requirements. They were instructed to complete training and practice trials as accurately as possible, and they received feedback about the accuracy of each action. During the main study, subjects were told to complete trials as quickly and accurately as possible, and they received no feedback. Pilot testing had previously determined the total number of tasks required to ensure that nurses were practiced users, the clarity and clinical meaningfulness of the assigned tasks, and technical computer issues such as network interference and consistency of screen drivers.

Requested Tasks

All participants were asked to perform two global tasks. First, they had to navigate about the interface to select the correct patient name and the nursing order application. Second, nurses interacted with the computerized nursing orders to complete the requested tasks. Four types of nursing orders represented typical interactions—create, activate (an order set), modify, and discontinue orders. Nursing orders are either single orders or order sets. Order sets are created for frequently used options as a collection of orders related to a diagnosis or common procedure. Identical tasks and order types were used in each block of trials for each interface. Five create, one discontinue, two modify, and two activate tasks were included in each block of trials. Eight tasks were single orders, and two tasks addressed order sets in each block of trials. The computer displayed requested tasks continuously throughout the trial in the upper right corner of the screen to prevent nurses from forgetting the particular request or the fictitious names of patients.

Nurse–Computer Interactions

Training and practice trials allowed nurses to rehearse typical requested tasks. Using a Latin square technique, the study was counterbalanced for interface types, task types, and blocks of trials to prevent progressive error or an order effect. Nurses completed a total of nine tasks prior to the main study tasks (n = 40) per interface. The computer program began timing nurses after they read the requested task and pressed the enter key or clicked "OK" to begin the trial. The timer stopped when nurses filed the completed order.

Nurses interacted with the computer by performing requested tasks, such as "Create a nursing order for
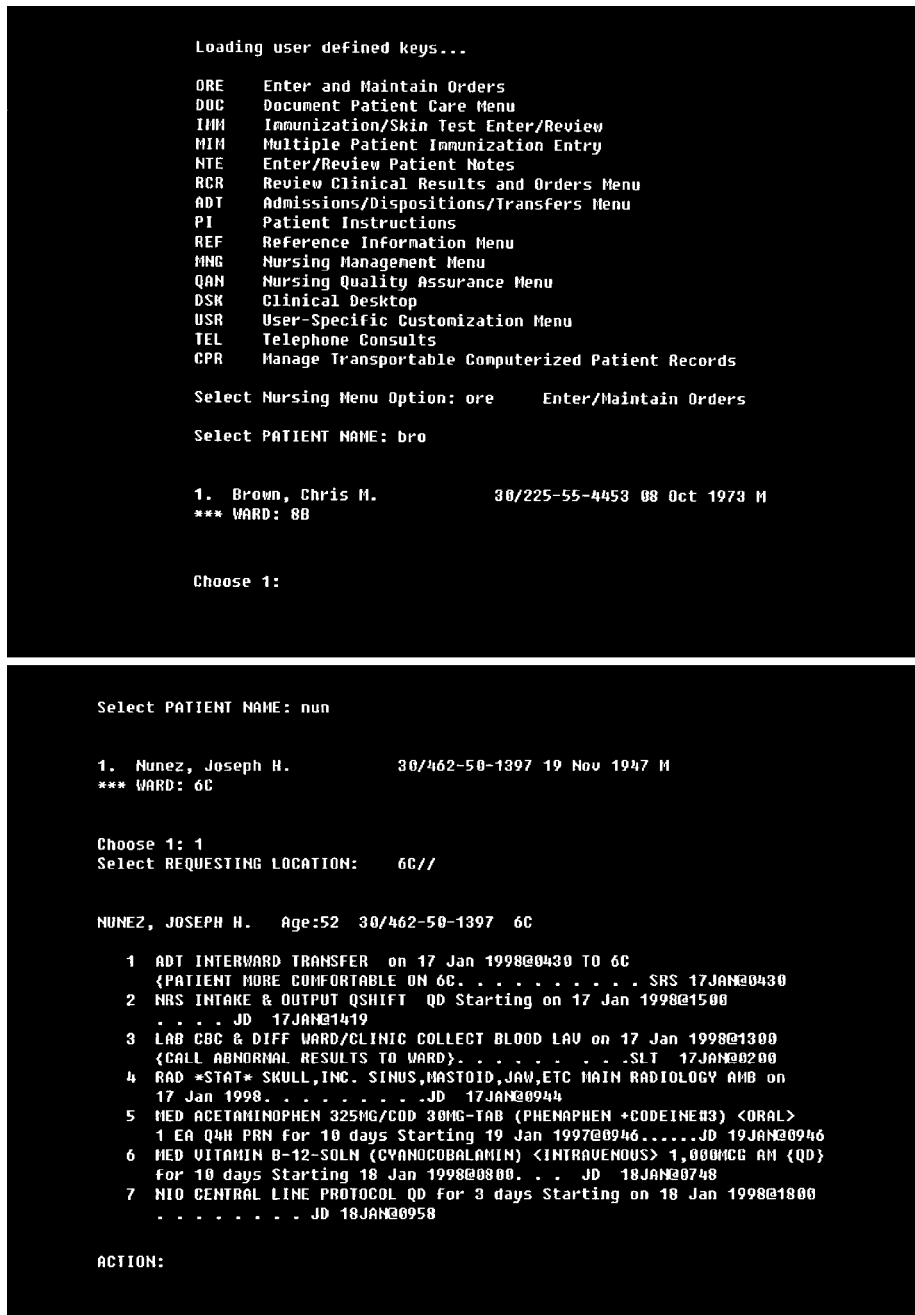
```
Loading user defined keys...

ORE     Enter and Maintain Orders
DOC     Document Patient Care Menu
IMM     Immunization/Skin Test Enter/Review
MIM     Multiple Patient Immunization Entry
NTE     Enter/Review Patient Notes
RCR     Review Clinical Results and Orders Menu
ADT     Admissions/Dispositions/Transfers Menu
PI      Patient Instructions
REF     Reference Information Menu
MNG     Nursing Management Menu
QAN     Nursing Quality Assurance Menu
DSK     Clinical Desktop
USR     User-Specific Customization Menu
TEL     Telephone Consults
CPR     Manage Transportable Computerized Patient Records

Select Nursing Menu Option: ore     Enter/Maintain Orders

Select PATIENT NAME: bro


1.  Brown, Chris M.          30/225-55-4453 08 Oct 1973 M
*** WARD: 8B


Choose 1:
```

```
Select PATIENT NAME: nun


1.  Nunez, Joseph H.         30/462-50-1397 19 Nov 1947 M
*** WARD: 6C


Choose 1: 1
Select REQUESTING LOCATION:     6C//


NUNEZ, JOSEPH H.   Age:52  30/462-50-1397  6C

    1  ADT INTERWARD TRANSFER  on 17 Jan 1998@0430 TO 6C
       {PATIENT MORE COMFORTABLE ON 6C. . . . . . . . . . SRS 17JAN@0430
    2  NRS INTAKE & OUTPUT QSHIFT  QD Starting on 17 Jan 1998@1500
       . . . . JD  17JAN@1419
    3  LAB CBC & DIFF WARD/CLINIC COLLECT BLOOD LAV on 17 Jan 1998@1300
       {CALL ABNORMAL RESULTS TO WARD}. . . . . . . . .SLT  17JAN@0200
    4  RAD *STAT* SKULL,INC. SINUS,MASTOID,JAW,ETC MAIN RADIOLOGY AMB on
       17 Jan 1998. . . . . . . .JD  17JAN@0944
    5  MED ACETAMINOPHEN 325MG/COD 30MG-TAB (PHENAPHEN +CODEINE#3) <ORAL>
       1 EA Q4H PRN for 10 days Starting 19 Jan 1997@0946......JD 19JAN@0946
    6  MED VITAMIN B-12-SOLN (CYANOCOBALAMIN) <INTRAVENOUS> 1,000MCG AM {QD}
       for 10 days Starting 18 Jan 1998@0800. . . JD  18JAN@0748
    7  NIO CENTRAL LINE PROTOCOL QD for 3 days Starting on 18 Jan 1998@1800
       . . . . . . . . JD 18JAN@0958


ACTION:
```

Figure 2 Text-based screens. *Top*, screen used to select a patient and navigate to an application. *Bottom*, screen used to interact with nursing orders. A user responds to a command line by typing, at the system prompt at the bottom of each screen, the first letter of the desired menu item. Patient names and details shown on the screens are fictitious.

vital signs q2h for Quincy Z. Brown on 3E'' or "Change the order for weigh daily to weigh every other day for Charlotte Smith on 7W." The nurses used primarily the keyboard with the text-based system and a mouse with the prototype GUI; however, the requested task types were identical for both interface styles.

Main Study Variables

Study variables were derived from the categories of the conceptual framework. Interface designs are com-

puter screen displays (computer actions) as defined by particular computer characteristics (programming, processing times, etc.). Response time, accuracy, and satisfaction are measured nurse behaviors that are a result of their characteristics and the requested tasks. Nurse characteristics were controlled by the study design, computer characteristics were controlled by the use of identical computers and by removal of the computers from the agency's network, and nurses received sufficient tasks in each display to be considered practiced users as is suggested by the trajectory in the framework.
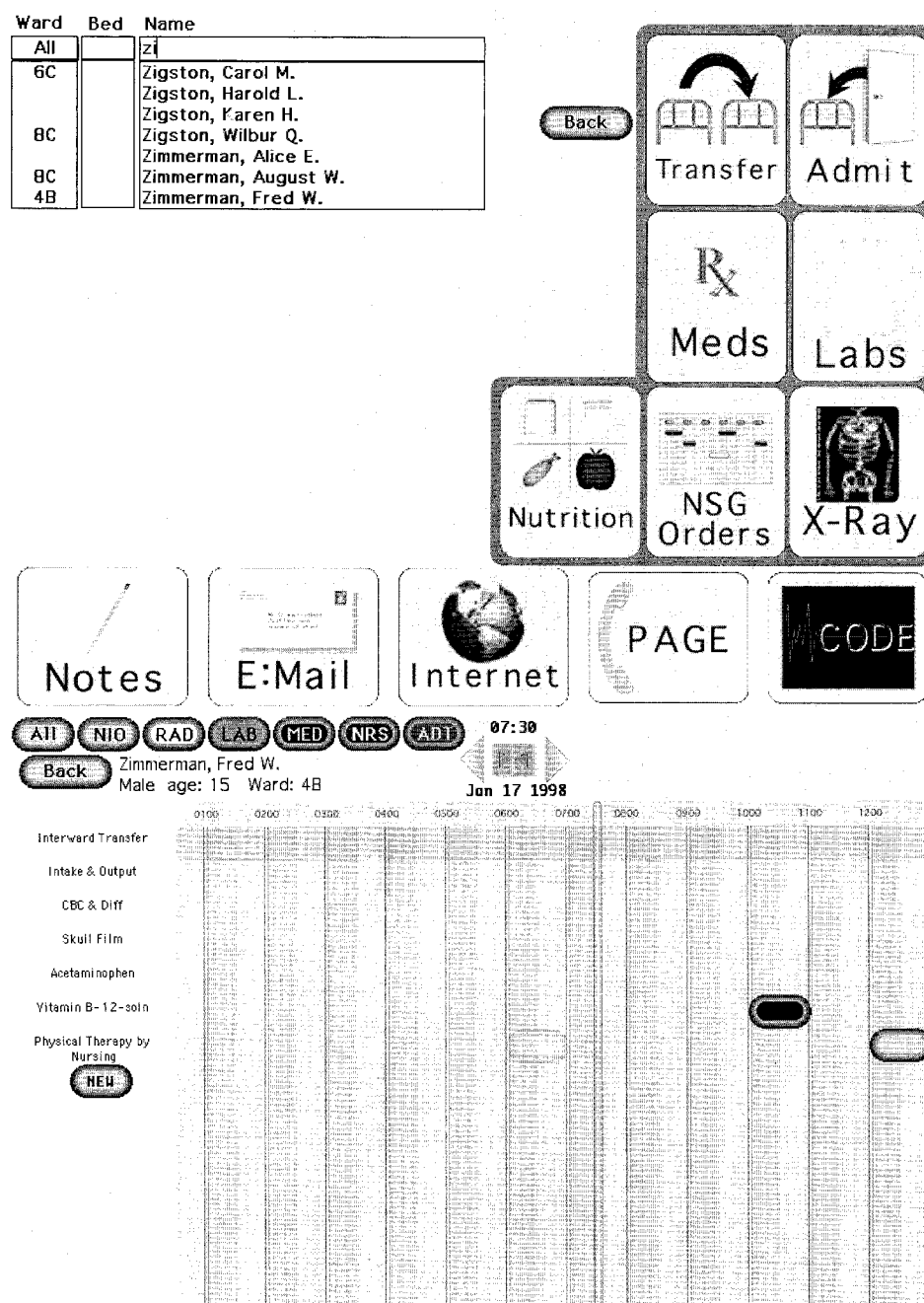
Figure 3 Graphical user interface (GUI) screens corresponding to the text-based screens shown in Figure 2. *Top,* screen used to select a patient name and navigate to the order application. *Bottom,* the redesigned order display corresponding to the order options menu on the text-based interface. Patient names and details shown on the screens are fictitious.

Computer Interface Design

The text-based interface (Figure 2) was a very close facsimile of that used in a Department of Defense hospital information system that is available worldwide. A facsimile of the text system was created on a desktop computer for several reasons. The interactions with the text facsimile were programmed to mirror the real system, to ensure that users would not detect differences between the legacy system and its facsimile. The actual system could not be used, because of the potential impact of system performance, including

polling by the network, variable system loads, and unexpected downtime.

As part of a larger study, qualitative research was completed to determine how nurses currently complete nursing orders with this text-based interface and how those interactions might be improved with a GUI. Using the results of that initial work and human factors principles, the text-based nursing orders were redesigned into a GUI prototype (Figure 3). By virtue of its design, a GUI can offer improved ways of viewing data, such as a time line of orders. However, to

Table 1 ■

Mean Response Times (RTs) and Error Rates with Use of the Graphical User Interface (GUI) and the Text-based Interface (N = 98)

	Response Time (sec)		Error Rate (per Task)	
	Overall	After Practice*	Overall	After Practice*
With GUI:				
Mean	32.51	29.49	0.36	0.29
SD	10.81	10.25	0.44	0.37
Range	18.25–87.44	17.99–81.83	0–6	0–1
With text-based interface:				
Mean	59.51	49.71	2.34	1.77
SD	16.13	17.99	2.11	1.65
Range	33.13–102.83	26.91–111.33	0–24	0–4

*Measured on performance for block 4.

ensure that the interface evaluations were equivalent, the authors tested only functions available in both interfaces.

Software Design

Both interfaces were developed using Macromedia Director, version 6.0. The GUI was developed in an iterative fashion by nursing focus groups and subject matter experts working in concert with human factors researchers and design engineers.

Response Time

A computer program automatically recorded the number of milliseconds it took nurses to complete each requested task. The mean times for practice, training, and each block of ten task trials were calculated for each interface.

Errors

The computer program also captured all keystrokes made during the completion of each task. Errors included accessing the wrong function or the wrong patient, typing in the wrong command, making typographic errors, and backtracking. The number of errors were summed per task and averaged across the ten trials in each block of requested tasks. The grand mean across the four blocks of trials was also calculated.

Accuracy

Besides summed errors, task accuracy was assessed. An entire task was counted as incorrect if it contained any errors. The total number of incorrect tasks was summed per block and averaged. The grand mean of incorrect tasks was calculated for the 40 tasks per display.

Satisfaction

Each participant's satisfaction was measured by use of the QUIS, or Questionnaire for User Interaction Satisfaction (short form).[29] Construct validity for the questionnaire was assessed by comparing a) software that was liked and disliked and b) a command-line versus menu-driven application. Total scores for the QUIS were significantly different for the liked and menu-driven applications. Inter-rater reliability using Cronbach's alpha was 0.94.[30]

Findings

Nurse–Computer Interaction Results

Response Time

As shown in Table 1 and Figure 4, mean performance response times across requested tasks began at a high
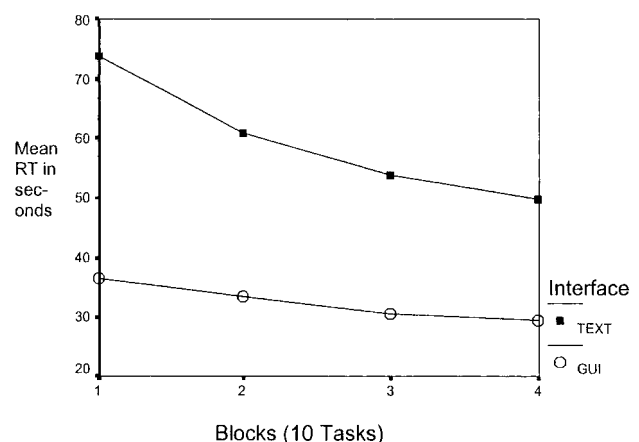

F i g u r e 4 Mean response times (RTs) across blocks of trials for the GUI and the text-based interface.

level in the first block of tasks for the text-based interface (73.77 sec) and dropped gradually by the fourth block of tasks (49.71 sec) as users became more proficient in the use of the system. The mean response time for the prototype GUI began at a much lower level (36.46 sec) and decreased progressively as the trials were completed (29.49 sec). By the last block of trials, nurses were still twice as fast using the GUI as using the text-based interface. Descriptive statistics for main variables of response time and errors are listed in Table 1.

The response-time data distribution here reflects a typical pattern in response-time studies (Figure 4) where skewness is expected. The response-time data distribution was positively skewed but at a low value (0.878). Stevens[31] indicated that significant skewness has only a slight effect, of a few hundredths, on the overall level of significance. Kurtosis, on the other hand, would be a more difficult issue.

Several methods are available to deal with significantly skewed data and outliers,[32] including data transformation, medians ANOVA analysis, and the use of cutoff response times. Ratcliff[32] compared these methods using simulations and found that none affected the alpha level, but the inverse transformation had higher power than the other methods. Researchers studying human–computer interaction typically do not transform data as is common in other arenas. They perform ANOVA using medians rather than means. The investigators chose not to report median testing for several reasons. First, the skewness of the data distribution was minor and was probably related to the large sample size for the study design. Second, power was not an issue in this study. Third, in five of the ten tasks in each block, the means and medians were equal (for the two activate, two modify, and one discontinue tasks). To ensure that the effect of the remaining five tasks did not influence the response-time results, the investigators ran repeated-measures ANOVA using both response-time means and medians. The prototype GUI was faster than the text-based interface on means ANOVA testing ($F(1,97) = 407.49$, $P < 0.0001$). Using medians ANOVA testing, the same comparison was $F(1,97) = 403.00$, $P < 0.0001$. The results were essentially the same using either analysis. Last, because the use of medians ANOVA testing is unusual outside human–computer interaction circles, it may be as controversial as is transformation of data within those circles. For these reasons, the means ANOVA analysis is reported here.

A three-factor (screen types/order types/and blocks of tasks) repeated-measures analysis was used on the data to determine any differences between the two types of interfaces. Each result for main effects was significant ($P < 0.001$) with the alpha level set at 0.05. As shown in Figure 4, the prototype GUI was significantly faster than the text-based interface ($F(1,97) = 407.49$, $P < 0.0001$). Simple main effects testing using the Tukey test revealed significantly faster response times for the prototype GUI than for the text-based interface for each block of requested tasks. The Bonferroni technique was used to adjust the P level to 0.01. A significant decrease in response time for both interfaces occurred from block 1 to block 4 ($F(1,97) = 75.17$, $P < 0.0001$), representing a significant learning effect for both interfaces. The main effect for order types was statistically significant ($F(1,97) = 119.43$, $P < 0.0001$). Simple main effects testing showed that the prototype GUI was significantly faster than the text-based interface for each block of trials where $P < 0.01$. However, of the four requested tasks, activating orders took nurses longer on both interfaces, averaging 57.75 sec. The mean time to complete other order types ranged from 35.67 to 43.01 sec.

The two-way interaction effects were statistically significant except for the block/order type interaction. Simple main effects testing using the Tukey test was done for each interaction effect. The Bonferroni technique was used to adjust the alpha level to 0.01. The interface and block interaction was statistically significant ($F(3,291) = 27.02$, $P < 0.001$). Response time was significantly faster for each block; however, response time with the text-based interface decreased an average of 20 sec across the four blocks, whereas the change with the GUI was much less dramatic. Difference in the patterns, one steep, one not, probably explain the finding, since at each level the prototype GUI was significantly faster.

The interaction for interface and order type was also statistically significant ($F(3,291) = 23.07$, $P < 0.001$). Completion of the requested task was significantly faster for each of the four order types with the prototype GUI (Figure 5); however, activating orders with the text-based interface took much longer than other order types. With either interface, activating an order set requires users to verify each order, contributing to the longer response time for this order type.

The three-way interaction of interface/order types/ blocks was significant ($F(9,873) = 3.13$, $P < 0.001$). Again, this result was probably driven by the steeper pattern of decrease in text response times across blocks and order types. There was no significant difference between the interaction response times for men and women for either interface.
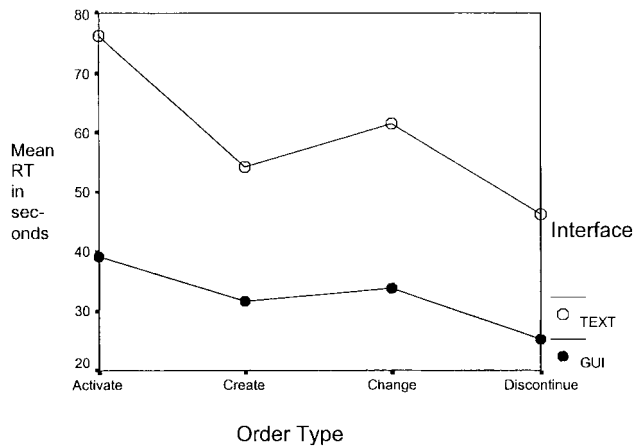
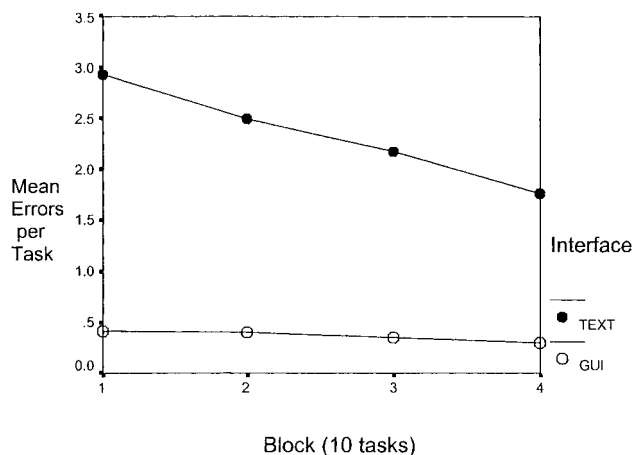**Figure 5** Response time interaction effect for interface and order type.

Figure 6 Mean errors for the GUI and the text-based interface.
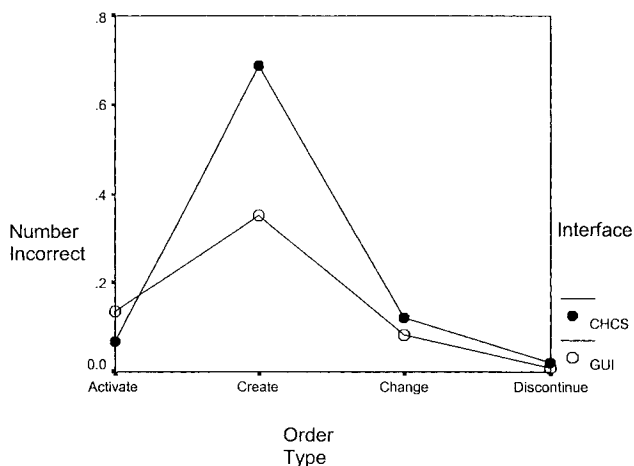

Figure 7 Incorrect tasks interaction effect for interface and order type.

Errors

Across requested tasks, the grand mean for errors was higher with the text-based interface than with the prototype GUI (Table 1). Errors with the GUI were lower throughout training and practice and throughout the four blocks of trials. In fact, the mean errors with the text-based interface were nearly six times higher even after subjects had completed all trials and were considered practiced users of the particular interface.

A three-factor (interface/order type/blocks of tasks) repeated-measures analysis was used to determine differences in error rates between the two interfaces. Overall, subjects committed significantly fewer errors with the prototype GUI than with the text-based interface ($F(1,97) = 192.1$, $P < 0.001$). Mean errors with the GUI began at a low level of 0.41 and decreased to 0.29 by the last block of tasks. Mean errors with the text-based interface, however, began at an average of 2.9 per trial in block 1 and decreased only to 1.77 by the last block.

Simple main effects testing revealed that the prototype GUI had fewer errors in each block of trials ($P < 0.001$). The average error rate decreased significantly over the four blocks ($F(1,97) = 7.50$, $P < 0.001$), representing a significant learning effect. Figure 6 shows the steeper decrease in mean errors with the text-based interface and probably explains this result. A significant main effect was also found for order type ($F(3,291) = 86.79$, $P < 0.001$). Additional analysis revealed that the prototype GUI had significantly fewer errors ($P < 0.001$) for each of the four order types.

The interaction effect for interface/block was significant ($F(3,291) = 5.06$, $P < 0.01$). This may be explained by the greater decrease in error rate across blocks during use of the text-based interface. That is, the prototype GUI errors were stable and fewer than text-based interface errors across trials, but text-based interface errors decreased more steeply across blocks; therefore, the difference between the number of text-based interface and GUI errors in block 1 was greater than the number in block 4.

The interaction effect for interface and order types was also statistically significant ($F(3,291) = 59.97$, $P = 0.001$). Interestingly, the text-based interface had higher error rates for all order types; however, the mean errors for creating orders with the text-based interface (5.0) was substantially higher than for the other types (all less than 2.0). The three-way and block/order types interaction effects were not statistically significant. The number of errors did not differ significantly across nursing specialties or experience with computers.

Table 2 ■

Mean Subjective Satisfaction Scores with Use of the Graphical User Interface (GUI) and the Text-based Interface (*N* = 98)

	Satisfaction Subscale				
	Learning	Screen Layout	Terminology	System Capabilities	Overall User Reaction
With GUI:					
Mean	7.58	7.81	7.71	7.57	7.59
SD	1.15	0.91	1.01	1.24	1.37
With text-based interface:					
Mean	4.59	5.34	5.50	4.24	3.81
SD	1.88	1.56	1.69	1.95	1.98

NOTE: Subscale scores could range from 1 (low satisfaction) to 8 (high satisfaction).

Accuracy

A corollary of total errors is the number of tasks completed incorrectly. A task was counted as incorrect if it contained any errors. Participants using the text-based interface completed more tasks incorrectly than those using the prototype GUI ($F(1,97)$ = 11.35, $P <$ 0.001). The mean number of orders completed incorrectly with the text-based interface was less than 2.4 percent in block 1 and dropped only slightly to stabilize just above 2 percent by block 4. The number of tasks completed incorrectly with the GUI was 1.6 percent in block 1 and dropped to 1.1 percent by block 4. By block 4, the number of incorrect tasks filed with the text-based interface was still twice that with the prototype GUI. Accuracy significantly improved across the blocks of tasks for both displays ($F(3,291)$ = 3.21, $P <$ 0.05). The main effect for order type was also significant ($F(3,291)$ = 145.00, $P <$ 0.0001). Simple main effects testing indicated that users had significantly greater numbers of inaccurate tasks with the text-based interface than with the prototype GUI for each order type ($P <$ 0.001), except for activating orders. The highest grand mean of incorrect tasks was in creating an order with the text-based interface (Figure 7), contributing substantially to the overall incorrect number of tasks. Without this order type, users averaged about a 1 percent incorrect task rate and the two system means were more closely approximated.

Satisfaction

Subjects rated the prototype GUI significantly higher by total mean QUIS scores and scores for each of the subscales—learning, screen layout, terminology, system capabilities, and overall user reaction. Table 2 summarizes these findings. Although the prototype GUI was consistently rated higher than the text-based interface, the differences in overall user reaction were more pronounced when participants used the GUI first. If they used the text-based interface first, partic-ipants assigned it a moderate mean score of 5.04. When they subsequently used the prototype GUI, they rated it significantly higher, at 7.18. ($P <$ 0.01). However, if participants used the prototype GUI first, they assigned it a higher mean score of 8.09. When they subsequently used the text-based interface, they rated it 2.5, which is significantly lower than the GUI rating and even lower than the initial text-based rating. Thus, had the study not been counterbalanced, an order effect would have influenced the results.

A one-factor (screen types) repeated-measures analysis was used on the data to determine any differences in users' satisfaction with the two interfaces. Each subscale factor, shown in Table 2, was significant. The prototype GUI was rated significantly higher than the text-based interface for learning ($F(1,97)$ = 15.46; $P <$ 0.0001), screen layout ($F(1,97)$ = 16.31; $P <$ 0.0001), terminology ($F(1,97)$ = 13.76; $P <$ 0.0001), system capabilities ($F(1,97)$ = 16.80; $P <$ 0.0001), and overall user reaction ($F(1,97)$ = 17.46; $P <$ 0.0001). Overall QUIS scores were significantly higher for the prototype GUI than for the text-based interface ($F(1,97)$ = 197.93; $P <$ 0.0001). Users felt strongly about the differences in usability. For the current text-based system, comments included, "Get rid of this ancient thing!" and "I'd rather stick needles in my eyes all day than use [the text-based system]." Interestingly, several expert text-based system users commented that they disliked the GUI because it slowed their performance; they had to remove their hands from the keyboard to manipulate the mouse.

Discussion

The military nurses who participated in this study had a low-moderate depth and breadth of experience with computers. This level of experience contrasts with the low level of computer experience of in a large

sample of civilian nurses in a tertiary care center on the East Coast early in the 1990s.[33] The experience of those in the current sample consisted primarily of word processing, e-mail, and a few clinical and hospital information system applications. Few participants had experience with nursing order entry. The operating room and psychiatric nurses reported lower levels of overall computer usage and knowledge. The lower levels may be related to the lack of computer equipment in operating rooms or the type of nurse–patient interactions in these areas. Intensive one-to-one or group interactions in inpatient psychiatric care and long periods of intraoperative care may provide fewer opportunities for nurses to achieve higher levels of computer experience. However, the lack of experience was not reflected in differences in participants' overall accuracy, speed, or error rates.

Overall, and for practiced tasks, nurses were faster using the prototype GUI than the text-based interface for computerized nursing orders. By the last block of tasks, these practiced users were still significantly faster using the prototype GUI, with nearly twice the response times using the text-based interface. The interaction effects were due primarily to the steeper decline of the text-based system over the blocks of trials. Although responses with the prototype GUI were faster for each order type, activating orders with the text-based interface consumed substantially longer mean times than did other order types. To activate orders using the text-based order sets, nurses must scroll through various orders before filing them. In the text-based system especially, this activity was very time-consuming. In addition, this result may reflect the more intuitive nature of the GUI display. The meaning represented by each of the displays made the system easier to learn and remember.

Findings about response times are consistent with those of several previous studies reporting the superiority of subjects' performance with GUI[16–22] and in contrast to ones reporting faster performance with text displays.[23–25] The results of this study align with those of the Temple group,[22] who claimed that GUI is superior for navigation tasks. However, the observed definitive differences here were not necessarily expected, because nurses currently used the text-based system in day-to-day activities and should have been more practiced in general commands on that system, although not in the specific commands of order entry.

The fact that clear differences continued despite the worldwide use of the text-based system may indicate that an even larger distinction in performance between the two interfaces might be observed among nurses who are completely naive users. Also, the dif-

ferences between the results of the current study and those of previous work may relate to the task types studied. Navigation and moderately complex tasks such as creating orders are not like the simple tasks in previous studies.

For practiced and overall tasks, nurses made significantly fewer errors using the prototype GUI than the text-based interface. The mean error rate for the text-based interface never approached the consistently low level of errors for the prototype GUI. Even after practice by the nurses, the mean error rate per task with the text-based interface was more than six times higher than with the prototype GUI.

The greatest mean number of errors was made in creating orders. The greater complexity of that task probably contributed to the higher error rate. Without that task, the percentage of incorrectly completed tasks was 1 percent or less across the other three order types. Interestingly, the order type that took the longest to complete, activation, did not produce as many errors or incorrect tasks. In fact, users of the text-based interface produced fewer mean incorrect tasks for this order type. Together, the results from total errors per task and incorrect tasks mean that while the error rate per task was high, nurses corrected most errors before filing. Still, incorrect tasks were filed more than 2 percent of the time with the text-based interface, primarily in creating orders. Because the text-based interface is currently installed worldwide in military hospitals, this error rate is disconcerting for creating orders in particular and for productivity concerns in general.

These findings are consistent with the results of the Temple group,[22] who found users completed approximately 90 percent of tasks correctly with a prototype GUI, compared with 68 percent with a text-based interface. However, the nurses in this study committed far fewer errors than did subjects in the Temple study.

Errors in task completion for word processing and spreadsheet tasks like those used by the Temple group have implications entirely different from those of errors in life-critical systems. Staggers[33] found a grand mean error rate of 8.75 percent when nurses detected clinical information on text screens of varying densities. That task was a simple information detection task, whereas this study used moderately complex tasks. The complex task here, especially for creating orders, probably contributes to the higher total error rate for the text system. Still, a 2 percent error rate for serious errors in critical systems could be unacceptable. While some errors are benign, the potential for more serious errors is great, especially with continued

use of the text-based system. An established error rate for health applications is not known, but even an optimally designed interface will not eliminate all human error.

This study analyzed error-inclusive data. To reduce variability, other studies such as those by Tullis[34] and Vincente et al.[35] used only error-free tasks for data analyses. Error-inclusive tasks are more realistic and generalizable to clinical situations and thus were included here.

The prototype GUI took substantially less time than the text-based interface to learn, meaning that nurses became practiced, proficient users with less practice. This finding has implications for improving nurses' productivity during training sessions and initial work with GUIs. It is clear that GUIs have the potential to substantially decrease the dedicated time required for both training and practice with clinical applications. The importance of this characteristic should not be underestimated, since most learning of systems comes from hands-on experience rather than from formal training. An interface that facilitates the learning processes and lessens the cognitive burden for health care providers should be further supported.

Nurses rated the prototype GUI significantly higher overall and on each subscale of the satisfaction measure. In fact, users felt strongly about the text-based system, as indicated by the informal remarks on their rating sheets. This result is consistent with those of Tomaiuolo,[16] who also found that users subjectively preferred a GUI. These results support the notion that the GUI has a higher usability rating by demonstrating enhanced efficiency, effectiveness, and satisfaction.[36] However, expert users of the text-based interface were frustrated by the extra hand movements required with GUIs. These power users need to be accommodated by the use of keyboard commands and shortcuts with GUIs.

The prototype GUI was developed through research into how nurses used the text-based interface and how those interactions might be improved. A similar process might have been used for improving the text-based interface alone. The two interfaces were compared on one platform to eliminate potential confounding variables. During actual use in clinical settings, differences in interfaces will be entangled with differences in other attributes, such as system architecture and complexity, system performance, network attributes, system loads, and response times for particular applications. Also, a limitation of this study is that the extended learning time for the text-based interface meant that nurses' errors may not have become completely stabilized by the end of the session. By the last block of trials, the performance curve should have been flat, as it was for response times with the prototype GUI. For this sample of nurses, the mean performance curve for errors with the text-based interface was not completely flat. If additional practice trials had been provided, the mean performance for errors might have been slightly better than that reported here. However, given the wide difference in means between the GUI prototype and text-based interface and the fact that errors typically only improved slowly after 49 tasks per interface, overall results would probably not have been affected.

Conclusions

Despite the increasing use of GUIs in health care agencies, this is one of the first empirical studies documenting differences between GUIs and text-based interfaces. The results of this study indicate that, for nursing orders, a GUI prototype significantly improved performance response times, error rates, and satisfaction ratings compared with an existing text-based interface for nurses across clinical areas within an enterprise system. Also, the GUI prototype was learned more quickly for navigation and order management tasks. These results have implications for program managers, designers, executives, and nurse computer users. The clear differences in these objective data for response times and error rates support consideration of the redesign of legacy systems to more modern user interfaces, especially for order entry, navigational tasks, and moderately complex ordering tasks. This redesign should include the use of human factors principles and input from user-centered focus groups. Nurse productivity and accuracy would both be enhanced. Also, the objective data provide a means of determining benefits and discriminating funding opportunities among applications for executives and computer application program managers.

Future work should include investigations into nurses' error rates, the classification of the seriousness of those errors, and the ways to design more error-free systems for health care providers, especially for creating orders. The full spectrum of order management represents a complex task; therefore, testing additional complex tasks would be beneficial. Designing and testing a nursing view of the computer-based patient record would provide valuable information. In fact, with the newness of usability assessments in health care, optimal human–computer interactions in health care promise to be a field of endeavor for years to come.

References ■

1. Davis FD. User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. Int J Man–Machine Studies. 1993;38:475–87.
2. Nygren E. (1997). From paper to computer screen: human information processing and interfaces to patient data. Int Med Inform Assoc Working Group 6:317–27.
3. Van Bemmel JH. Medical data, information, and knowledge. Methods Inf Med. 1988;27:109–10.
4. Lindberg DA. Interview: Donald A. B. Lindberg, medicine-computer alchemist. Gov Comput News. Mar 29, 1993:17.
5. Orthner H. Series preface. In: Saba VK, McCormick KA. Essentials of Computers for Nurses. New York: McGraw-Hill, 1996.
6. Ozbolt J. Research priorities in nursing informatics: panel presentation. Proc 16th Annu Symp Comput Appl Med Care. 1992.
7. Bearman M, McPhee W, Cesnik B. Designing interfaces for medical information management systems. In: Greenes RA, Peterson HE, Protti DJ (eds). Proceedings of the 8th World Congress on Medical Informatics; Vancouver, Canada; Jul 23–27, 1995.
8. Staggers N. Usability concepts for the clinical workstation. In: Ball M, Hannah K, Newbold S, Douglas M (eds). Nursing Informatics: Where Caring and Technology Meet. 2nd ed. New York: Springer-Verlag, 1995:188–99.
9. Staggers N. Human factors: the missing element in computer technology. Comput Nurs. 1995;9(2):47–9.
10. Claus P, Gibbons P, Kaihoi B, Mathiowetz M. Usability lab: a new tool for process analysis. HIMSS Proc. 1997;2:150–9.
11. Sittig DF, Yungton J, Kuperman GJ, Teich JM. A graphical user interaction model for integrating complex clinical applications: a pilot study. Proc AMIA Annu Symp. 1998:708–12.
12. Patel VL, Kushniruk AW. Interface design for health care environments: the role of cognitive science. Proc AMIA Annu Symp. 1998:29–37.
13. Dix A, Finlay J, Abowd G, Beale R. Human–Computer Interaction. London, England: Prentice Hall Europe, 1998: 427–35.
14. Head AJ. Question of interface design: how do online service GUIs measure up? Online. 1997, May–Jun 1997:20–9.
15. Murphy T. Innovative interfaces: new GUIs help zero in on information. Inf Highways. 1997;4(4):20–3.
16. Tomaiuolo NG. End-user perceptions of a graphical user interface for bibliographic searching: an exploratory study. Comput Libr. Jan 1996:33–8.
17. Michard A. Graphical presentation of Boolean expressions in a database query language: design notes and an ergonomic evaluation. Behav Inf Technol. 1982:70–9.
18. Rohr G. Understanding visual symbols. IEEE Computer So-

ciety Workshop on Visual Language. Hiroshima, Japan: The Computer Society Press, 1984:181–91.
19. Woodgate H. The use of graphical symbolic commands (icons) in application programs. In: Vandon: CE (ed). Proceedings of the European Graphics Conference and Exhibition. Amsterdam, The Netherlands: Elsevier, North Holland, 1985:25–36.
20. Rauterberg M. An empirical comparison of menu-selection (CUI) and desktop (GUI) computer programs carried out by beginners and experts. Behav Inf Technol. 1992;11:227–36.
21. Davis S, Bostrom R. An experimental investigation of the roles of the computer interface and individual characteristics in the learning of computer systems. Int J Human–Computer Interaction. 1992;4(2):143–72.
22. Temple, Barker & Sloane, Inc. The Benefits of the Graphical User Interface: A Report on New Primary Research. Redmond, Wash: Microsoft Corp., 1990.
23. Whiteside J, Jones S, Levy PS, Wixon D. User performance with command, menu, and iconic interfaces. Proceedings of the CHI '85 Conference: Human Factors in Computing Systems. Reading, Mass: Addison-Wesley, 1985:185–91.
24. D'Ydewalle G, Leemans J, Van Rensbergen J. Graphical versus character-based word processors: an analysis of user performance. Behav Inf Technol. 1995;14(4):208–14.
25. Carroll J, Mazur S. Lisa learning. IEEE Comput. 1986;19: 35–49.
26. Staggers N, Parks P. A framework for research on nurse–computer interactions: initial applications. Comput Nurs. 1993;11(6):282–90.
27. Staggers N, Parks P. Collaboration between unlikely disciplines in the creation of a conceptual framework for nurse–computer interactions. Proc 16th Annu Symp Comput Appl Med Care. 1993:661–5.
28. Staggers N. The Staggers nursing computer experience questionnaire. Appl Nurs Res. 1994;7(2):97–106.
29. Norman KL, Shneiderman BA, Harper BD, Slaughter LA. Questionnaire for User Interaction Satisfaction, Version 7.0: Users' Guide, Questionnaire and Related Papers. College Park, Md: University of Maryland, 1998.
30. Chin JP, Diehl VA, Norman KL. Development of an instrument measuring user satisfaction with the human–computer interface. Proceedings of the CHI '88 Conference: Human Factors in Computing Systems. Reading, Mass: Addison-Wesley, 1988: 213–8.
31. Stevens J. Applied Multivariate Statistics for the Social Sciences. 3rd ed. Mahwah, NJ: Lawrence Erlbaum, 1996:242–3.
32. Ratcliff R. Methods for dealing with RT outliers. Psychol Bull. 1993; 114:510–32.
33. Staggers N. The impact of screen density on clinical nurses' computer task performance and subjective screen satisfaction. Int J Man–Machine Studies. 1993;39:775–92.
34. Tullis T. Predicting the Usability of Alphanumeric Displays. [PhD thesis]. Houston, Tex: Rice University, 1984.
35. Vincente K, Hayes B, Williges R. Assaying and isolating individual differences in searching a hierarchical file system. Human Factors. 1987;29:349–59.
36. Dumas JS, Redish JC. A Practical Guide to Usability Testing. Norwood, NJ: Ablex, 1993.