Assist Technol 1994;6:42–53

Learning and Performance of Able-Bodied Individuals Using Scanning Systems with and without Word Prediction

Heidi Horstmann Koester, M.S., and Simon P. Levine, Ph.D.

Rehabilitation Engineering Program, Department of Physical Medicine and Rehabilitation, University of Michigan, Ann Arbor, Michigan

This study examines how the cognitive and perceptual loads introduced by a word prediction feature impact learning and performance. Two groups of able-bodied subjects transcribed text using two row–column scanning systems for 10 consecutive trials each. The two systems differed only in that one system had a word prediction feature. Subject groups differed in their order of system use. The results show that, under the conditions of this study, the word prediction system was not substantially more difficult to learn, but it did not yield a statistically significant improvement in text generation rate. This suggests that the cost of using this word prediction system balanced the benefit of the keystroke savings achieved by these subjects. The relationship between keystroke savings, cost in item selection rate, and improvement in text generation rate is explored in order to provide insight into this outcome.

Key Words: Assistive technology—Augmentative communication—Rate enhancement—Word prediction—User performance modeling.

A wide range of assistive technology systems has been developed to facilitate function in a variety of areas, including powered mobility, environmental control, augmentative communication, and computer access. All of these systems include a user interface, which accepts some type of user input to control the system in the desired fashion. In many instances, the interface is designed with a primary focus on utilizing the motor abilities of the intended user as efficiently as possible.

While the goal of improving motor efficiency is an important one, it has been recognized that this may also place increased cognitive and perceptual requirements on the user, leading to unknown effects on the user's ability to learn and use the system. This dilemma exists in almost every area of assistive technology (1), but it has been discussed most frequently in connection with computer access and augmentative communication (AAC) systems, especially those that employ a rate enhancement feature such as word abbreviations (2), message encoding (3,4), or word prediction (5–7).

This paper focuses on the trade-off between improved motor efficiency and increased cognitive-perceptual loads in the context of word prediction systems. These systems attempt to predict the word intended by the user by presenting the user with a set of word choices. Word prediction choices are typically displayed in a short list and refined as the user selects additional letters. Because many words can be completed by choosing from the list rather than single letter spelling, the number of selections required per word can be substantially reduced.

The motor efficiency of word prediction systems is often measured in percentage of keystrokes saved.[1] Experimental measurements on two different prediction systems show a range of 37–47% keystroke savings over several different types of text samples (8). Clinical data on actual users reveal a broader range of 23–58% keystroke savings (9–11). Many of the clinical reports are anecdotal, with little spe-

Address correspondence and reprint requests to Heidi Horstmann Koester, 1C335 University Hospital, Ann Arbor, MI 48109-0032 U.S.A.

[1] "Keystrokes" are broadly defined to include key presses in a direct selection system, as well as items selected in other ways, such as through scanning or Morse code.

cific information on the conditions under which keystroke savings were measured, but most of them are consistent with Higginbotham's experimentally determined range of 37–47% (8).

While word prediction systems can be successful in reducing the motor requirements for text generation, this alone does not always yield a significant improvement in rate. Figure 1 shows a scatter plot of improvements in text generation rate at different levels of keystroke savings for eight single case reports (9–12). While some users enjoyed substantial improvement relative to letter-by-letter spelling, others improved only marginally or even decreased in speed. The most dramatic example of this is seen in Newell et al. (10), in which one subject's rate doubled (for an improvement of 100%), with a keysaving of 58%, while another subject's rate decreased by approximately 5%, despite an average keysaving of 53%. Additionally, at least two clinical case studies report that while efficiency may improve substantially, text generation rate may not (13,14).

These data support the long-standing hypothesis that using a predictive system to decrease the number of necessary selections may increase the time required to make each selection, leading to unknown effects on overall performance (4,5,7,15). One way of conceptualizing this trade-off is through the performance model developed by Rosen and Goodenough-Trepagnier (16). In this model, the average time per word, τ, is expressed as $\tau = C*L*T$, where C is the linguistic cost, or the average number of selection units per word, L is the average number of acts required per selection unit, and T is the average time per act. When word prediction is added to a letter-by-letter spelling system, the keystroke savings yields a decrease in C, while L is unchanged. However, T may increase because the act of making a selection has increased in complexity. The net impact on τ, the average time per word, depends on the relative magnitude of these changes in C and T.

Empirical information on just how much, and under what conditions, the time per selection may increase has not been reported in previous studies of performance with word prediction. However, several investigators have attempted to estimate the cost of using word prediction by analyzing the component processes involved, such as searching the word list, or deciding whether to search in the first place (6,7,12,17). One such analysis estimated the extra time per selection to be 1.22 seconds (7). These analyses help explain why time per selection may increase when using word prediction, but their quantitative accuracy has not been verified.

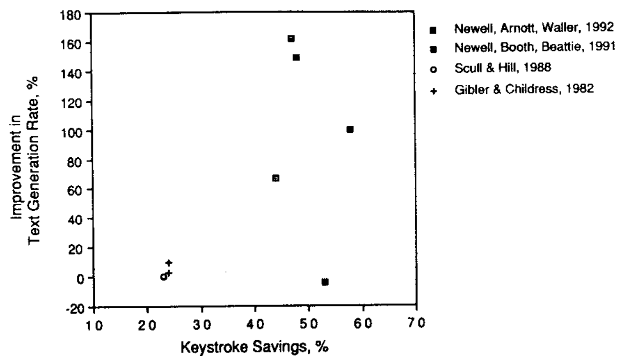

FIG. 1. Improvements in text generation rate reported for various levels of keystroke savings. Each point in the scatter plot corresponds to the performance of a single individual. Note the lack of a clear relationship between rate improvement and keystroke savings.

RESEARCH QUESTIONS

The general goal of our research is to improve understanding of the trade-off between increased cognitive-perceptual requirements and decreased motor loads in assistive technology systems, with a current focus on word prediction systems. In particular, we would like to eventually define the conditions under which word prediction improves text generation rate and those under which it does not. These conditions involve characteristics of the user, the specific implementation of the system, and the particular way in which the user employs the system. Ultimately this understanding may provide a means of simulating the effect of different conditions on overall performance, which would be a potentially powerful tool for designers as well as clinicians (6,7).

This paper reports on recent empirical and theoretical progress toward these goals. The empirical data come from a study in which able-bodied subjects used scanning systems with and without word prediction. Theoretical concepts are presented to explore the underlying reasons behind the results obtained.

METHODS

Subjects

Six able-bodied subjects were employed. All subjects were graduate students who had no cognitive, perceptual, or linguistic impairments. Each had some conceptual familiarity with assistive technology, but none of the subjects had direct prior experience with the systems studied.

_	E	A	R	D	U
T	N	S	F	W	B
O	H	C	P	V	J
I	M	Y	K	Q	,
L	G	X	Z	.	'
<	r	q			

FIG. 2. Arrangement of letter items in the Letters-only system. Letters are arranged in order of frequency of occurrence in English, with the upper left corner as the most frequently chosen character. The three special items on the bottom row correspond to backspace, carriage return, and quit, respectively.

_	E	A	R	D	U	1: the
T	N	S	F	W	B	2: to
O	H	C	P	V	J	3: is
I	M	Y	K	Q	,	4: that
L	G	X	Z	.	'	5: of
<	bw	r	q			6: you

FIG. 3. Arrangement of letter and word items in the Letters + WP system. The letter matrix is identical to that of the Letters-only system. The four special items on the bottom row per form are for backspace, list selection correction, carriage return, and quit, respectively.

Interfaces

The two interfaces under study were developed specifically for research purposes, to gain sufficient control over the system configuration as well as the means of data collection. Both interfaces used single switch row–column scanning as the basic input method. The first interface, referred to as "Letters-only," required letter-by-letter spelling, using a fixed, frequency-based, letter matrix, shown in Figure 2. Scanning proceeded continuously row-by-row until the switch was pressed to choose a particular row; each column in that row was then scanned until the switch was pressed a second time to choose the desired item. Scanning then resumed from the top row. The scan speed (i.e., the length of time that a row or column remained highlighted) was fully adjustable, as was an extra "row delay" for the first row and column in the matrix. These parameters were initially set at 750 milliseconds for the scan speed and 250 milliseconds for the row delay.

The second interface, referred to as "Letters + WP," used the same letter matrix augmented by a word prediction feature, shown in Figure 3. Characteristics of the word prediction feature included a six-word list, fixed prediction dictionary, and fixed order of words in the list. The method of item selection was similar to that of the Letters-only system, except that a "half-and-half" scanning pattern was used, in which scanning first alternated between the letter matrix and the word list. With this method, the first switch hit chose the desired half. If the matrix was chosen, two more switch hits were required to choose a letter. If the word list was chosen, one more switch hit was required to choose a word. After an item was selected, scanning resumed on the matrix half. The timing parameters were fully adjustable; these included the scan speed, the row delay, and an extra delay on the matrix half. (Note that the extra pause on the matrix half was only added on the first scanning cycle.) These parameters were initially set at 750 milliseconds for the scan speed, 250 milliseconds for the row delay, and 500 milliseconds for the half delay.

Experimental Design

The six subjects were evenly divided into two groups. The order of system use was as follows:

Group A

1. Letters-only, training session
2. Letters-only, test sessions
3. Letters + WP, training session
4. Letters + WP, test sessions

Group B

1. Letters + WP, training session
2. Letters + WP, test sessions
3. Letters-only, training session
4. Letters-only, test sessions

Training required one session and combined verbal instruction and practice. Subjects were given the goal of achieving their maximum possible rate with each interface while keeping to less than 10% timing errors (i.e., item was correct but not selected at first opportunity) and less than 5% incorrect selections (i.e., both corrected and uncorrected errors). During Letters + WP training, the rationale behind word prediction was explained, but subjects were not given specific guidelines or strategies for when to use the feature. Subjects practiced using the system at the relatively slow initial speed on a text sample until they could select text with 95% accuracy. This criterion was generally reached within two sentences.

Testing sessions occurred twice a week and involved two text transcription tests, each preceded by a warm-up period. All subjects began the study with the initial timing parameters. Before the first test, they had three 2-minute warm-up periods in which to tune the timing parameters to match their skill, with an experimenter available to provide assistance as needed. Parameters could be adjusted in increments of 25, 50, or 75 milliseconds, as long as errors did not exceed the established error criteria. Timing errors were detected and tracked in real time by the software, while selection errors were noted by the experimenter during each session. The limitation on errors helped ensure that the timing parameters were neither too fast nor too slow for each subject. Subsequent tests contained a single warm-up period, and parameters could be modified before or after the warm-up under these same guidelines.

Subjects transcribed 10 unique five-sentence blocks of text for each system. Text blocks were drawn from published typing tests, already matched with respect to average word length, average syllables per word, and percentage of words with a high frequency of occurrence (18). Slight revisions to these tests were made to match overall length and average scan steps across blocks.[2] Individual text blocks were not precisely matched with respect to the word prediction characteristics (e.g., keystroke savings, proportion of words in the dictionary); the specific values for these at each transcription trial are provided in Table 1. Overall, the dictionary contained 83% of the words in the Letters + WP text blocks, providing an average keystroke savings of 42%.

Subjects read the sentences from index cards, containing one sentence per card. They had 20 seconds to flip to a card and read the sentence. During this period, no selections could be made. Scanning resumed automatically at the end of the "freeze" period, and subjects then transcribed the sentence using the assigned interface. Errors could be corrected by selecting special items for backspacing single letters as well as word list selections. The sentence card remained in view for reference throughout transcription.

[2] The number of scan steps associated with an item on the letter matrix is the number of rows and columns that must be scanned to reach the item.

TABLE 1. Characteristics of text blocks used in Letters + WP trials

Letters + WP trial	Keystroke savings (%)	Prediction success (%)
1	42	79
2	36	70
3	44	81
4	41	84
5	45	85
6	43	89
7	45	90
8	38	77
9	47	92
10	40	79

Data Analysis

All items selected by subjects were timed and stored by the software in real time. The configuration parameters used during a session were also recorded with the item data. An experimenter was present throughout each session to record observations of subject behavior.

The two primary dependent measures were text generation rate and item selection rate. Text generation rate was measured in characters per minute (cpm), to be independent of word length, and incorporated all characters generated in a trial, including punctuation, timing errors, selection errors, and error corrections. However, the trends described below remain consistent whether or not these factors are included in the text generation rates. Item selection rate for each trial was defined as the number of items that were selected per unit time. Note that for the Letters-only system, the text generation rate and item selection rate were necessarily identical, since each item selection generated only one character. For the Letters + WP system, the text generation rate was necessarily faster than the item selection rate, because each selection generated more than one character, on average, due to use of the word list.

Statistical differences in these rates across groups, systems and trials were determined using a repeated measures ANOVA model, with group or system as the between-subjects factor, and trial as the repeated measures (within-subjects) factor. Interactions between the system and trial factors were examined to identify differences in learning rates.

RESULTS

Analysis of Text Generation Rate

Figure 4 shows the average text generation rate achieved by each group over the 20 trials. Across

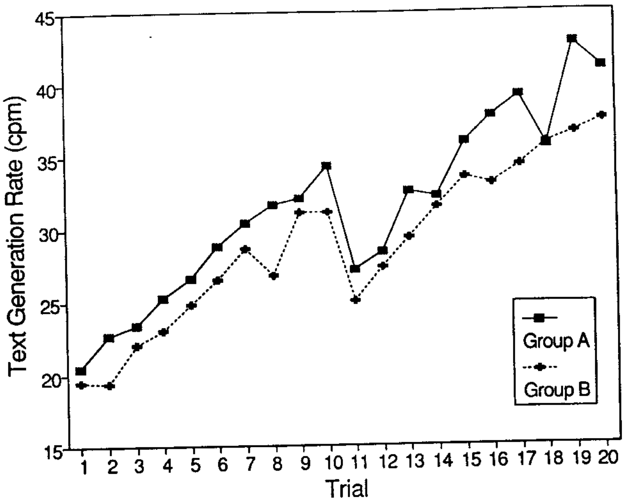

FIG. 4. Average text generation rates achieved by each group at each trial. The discontinuity at trial 10 corresponds to the point at which groups switched to a different system.

all trials, Group A is somewhat faster than Group B. The results can be examined more closely by considering the two halves of the study: the first half where Group A used Letters-only and Group B used Letters + WP for 10 trials each, and the second half where groups switched systems.

Over the 10 trials of the first half, the Letters + WP subjects made an average of 34.8% fewer selections than Letters-only subjects, yet their text generation rate was an average of 8.2% slower. The left half of Figure 4 shows this difference graphically, while Table 2 provides the average and standard deviation for subjects' rates at each trial. The 8% rate difference was not statistically significant, based on an ANOVA test for main effect of system, with repeated measures on trials ($p = 0.442$). From the first to tenth trials, the average rate for subjects using Letters-only ranged from 20.4 cpm (3.6 wpm) to 34.3 cpm (6.0 wpm), with the fastest user achieving 38.0 cpm (6.7 wpm).[3] For subjects using Letters + WP, this range was 19.5 cpm (3.4 wpm) to 31.1 cpm (5.5 wpm), with the fastest user attaining 35.7 cpm (6.3 wpm).

Over the second half of the experiment, subjects who used Letters + WP (Group A) made an average of 36.4% fewer selections and had a text generation rate that was an average of 8.7% faster than those who finished with Letters-only (Group B). Table 2 shows the average and standard deviation for subjects' rates over the second half, and the right half of Figure 4 illustrates the differences graphically.

[3] Words per minute calculations assume 5.7 letters per word.

TABLE 2. Text generation rate data for each trial

Group	Trial	First half $\bar{x}$	First half SD	Second half $\bar{x}$	Second half SD
A	1	20.42	2.70	27.12	3.33
	2	22.68	3.57	28.39	3.50
	3	23.41	4.71	32.50	3.97
	4	25.22	4.40	32.12	2.78
	5	26.62	5.35	35.94	3.57
	6	28.82	5.15	37.66	4.84
	7	30.38	4.55	39.07	2.57
	8	31.66	4.45	35.64	3.37
	9	32.11	3.86	42.76	4.54
	10	34.31	3.40	41.06	4.72
B	1	19.47	1.76	24.97	2.89
	2	19.42	0.86	27.27	3.87
	3	22.05	1.53	29.34	4.47
	4	23.03	1.26	31.45	3.90
	5	24.82	1.84	33.46	6.05
	6	26.53	3.19	33.08	5.00
	7	28.59	2.90	34.30	4.65
	8	26.78	3.31	35.82	5.57
	9	31.09	3.97	36.57	3.44
	10	31.07	4.37	37.43	2.12

Means and standard deviations are in characters per minute.

As in the first half, an ANOVA test for main effect of system, with repeated measures for trials, showed that this difference was not statistically significant ($p = 0.418$). From the first to tenth trials, the average rate for subjects using Letters-only ranged from 25.0 cpm (4.4 wpm) to 37.4 cpm (6.6 wpm), with the fastest user at 42.3 cpm (7.4 wpm). For subjects using Letters + WP, this range was 27.1 cpm (4.8 wpm) to 41.1 cpm (7.2 wpm), with the fastest subject achieving 48.0 cpm (8.4 wpm).

There did not appear to be a notable difference in the learning rates between systems in either half of the study. In the first half, average text generation rate improved by 59.6% for Letters + WP subjects and 68.0% for Letters-only subjects. For both systems, then, practice was a major factor in determining text generation rate, and the repeated measures ANOVA showed the effect of trial to be highly significant ($p < 0.001$). Graphically, the rates of improvement for both systems look similar (Fig. 4), which is statistically supported by the lack of a significant interaction between trial and system ($p = 0.553$).

In the second half, practice also significantly improved text generation rate for both groups ($p < 0.001$), with rate improving by 51.4% for Letters + WP subjects and 49.9% for Letters-only subjects. Statistically, the repeated measures ANOVA did

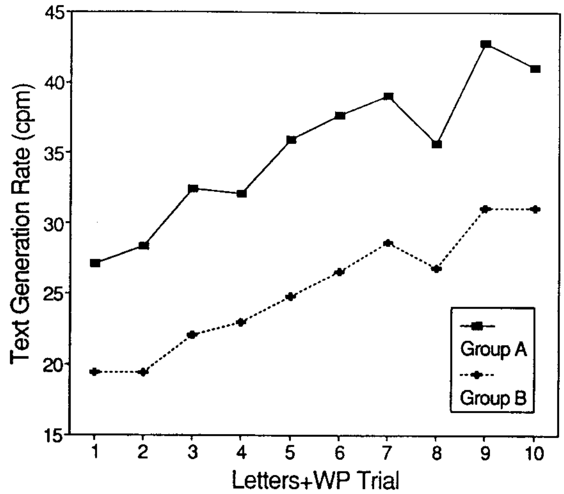

FIG. 5. Average text generation rates achieved at each trial, when each group used the Letters + WP system. Note that this shows Group A's performance for the second half of the study and Group B's performance for the first half.

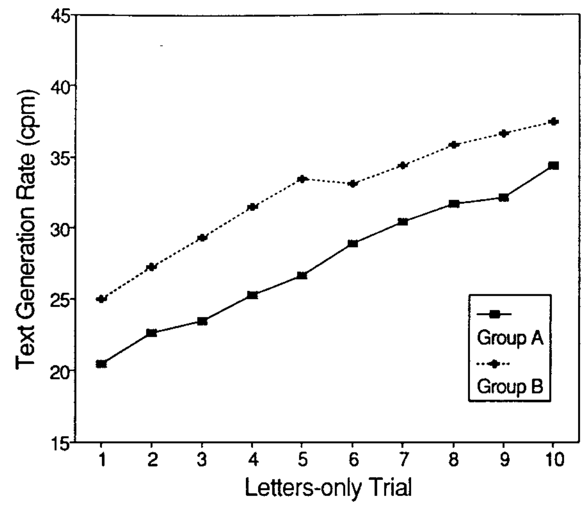

FIG. 6. Average text generation rates achieved at each trial, when each group used the Letters-only system. Note that this shows Group A's performance for the first half of the study and Group B's performance for the second half.

show a significant interaction between trial and system ($p = 0.005$). This may indicate some difference in learning rates, since the Letters-only group improved more slowly in trials 6–10 than in trials 1–5. However, a clear comparison to the improvement of the Letters + WP group is difficult to make due to the variation in the keystroke savings offered in each text block (see Table 1).

A second set of analyses compared the groups' performances when they used the same system, in an attempt to understand how prior experience with one system might affect subsequent performance on the other system. For example, in comparing the performance of both groups on the Letters + WP system, Group A has already had 10 trials of experience with Letters-only, while Group B has had no prior experience. Therefore, it is not surprising that when Group A subjects switched to Letters + WP, their performance with that system was an average of 39.9% faster than that of Group B (Fig. 5). An ANOVA test for main effect of group, with repeated measures on trials, showed this difference to be significant, at $p = 0.015$.

This experience effect was not quite so pronounced, however, when groups were compared using Letters-only. In this case, Group B subjects have had 10 trials of prior experience with the Letters + WP system, while Group A subjects have had none. When Group B subjects switched to Letters-only, their performance was an average of 18.2% better than that of Group A (Fig. 6), which was not statistically significant ($p = 0.224$). The gap between groups began to close slightly during the last five trials, ending at a 9.1% difference for trial 10.

In both of these "same system" comparisons, 10 trials of experience with one system improved the subsequent performance with the other system. Learning rate, however, did not seem to be significantly affected by prior experience. Graphically, the slopes of the lines in each of Figures 5 and 6 are roughly the same, suggesting that the learning rates for groups were similar when each used the same system. Furthermore, neither of the ANOVA tests for differences between groups showed a significant interaction between trial and group ($p = 0.428$ for Letters + WP; $p = 0.463$ for Letters-only).

Analysis of Item Selection Rate

Analysis of the item selection rate data provides some insight into why the use of word prediction did not provide a significant enhancement of text generation rate. For both halves of the study, the item selection rate for Letters + WP users was significantly slower than that of Letters-only users, as shown graphically in Figure 7 and numerically in Table 3. In the first half, the selection rate for Letters + WP users (Group B) was 40.5% slower than that of Letters-only users (Group A). This difference was analyzed using an ANOVA for main effect of system, with repeated measures on trials, and found to be significant ($p = 0.011$). The second half of the study also showed a large difference between selection rates for the two systems, since the rate for Letters + WP selections was an average of 30.9% slower than Letters-only selections (significant at $p = 0.029$).

This decrease in item selection rate means that

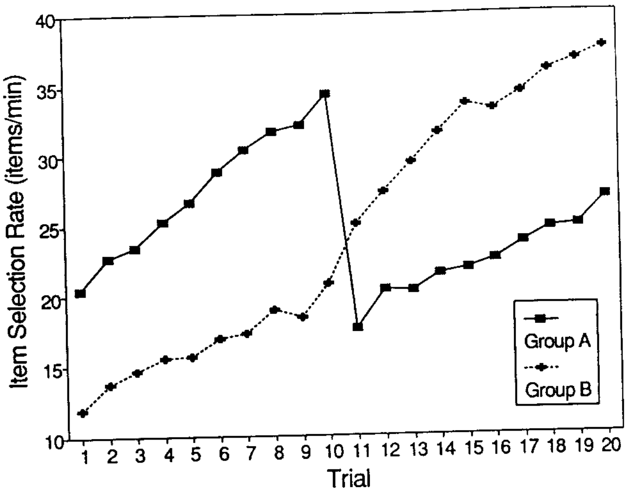

FIG. 7. Average item selection rates achieved by each group at each trial. The discontinuity at trial 10 corresponds to the point at which groups switched to a different system.

TABLE 3. Item selection rate data for each trial

Group	Trial	First half $\bar{x}$	First half SD	Second half $\bar{x}$	Second half SD
A	1	20.42	2.70	17.52	2.90
	2	22.68	3.57	20.33	3.55
	3	23.41	4.71	20.25	3.38
	4	25.22	4.40	21.43	3.12
	5	26.62	5.35	21.73	3.04
	6	28.82	5.15	22.41	3.51
	7	30.38	4.55	23.56	2.80
	8	31.66	4.45	24.59	3.34
	9	32.11	3.86	24.79	3.30
	10	34.31	3.40	26.82	3.96
B	1	11.93	1.75	24.97	2.89
	2	13.71	1.36	27.27	3.87
	3	14.62	1.32	29.34	4.47
	4	15.56	1.18	31.45	3.90
	5	15.63	1.54	33.46	6.05
	6	16.90	1.89	33.08	5.00
	7	17.32	2.07	34.30	4.65
	8	18.90	1.27	35.82	5.57
	9	18.38	0.94	36.57	3.44
	10	20.72	1.75	37.43	2.12

Means and standard deviations are in items per minute.

the time required to make each selection was consistently longer for Letters + WP users than for Letters-only users. In the first half of the study, each selection made by the Letters + WP users took an average of 1.5 seconds longer than each Letters-only selection, starting at 2.1 extra seconds in the first trial and decreasing to 1.2 seconds by the tenth trial. The second half showed a similar picture; each selection in Letters + WP took an average of 0.9 seconds longer than a Letters-only selection, starting at 1.1 extra seconds in the first trial and decreasing to 0.7 extra seconds by the tenth trial.

DISCUSSION

In order to interpret these results, it is important to consider the potential impact that specific features of the experimental design may have had on the outcome. Theoretical principles are also discussed, to place the results in a broader context and to define avenues for future research.

Learning

No statistically significant differences in either the achievement of basic competence or development of expertise were observed between systems over the course of this study. All subjects were able to select items accurately within a single training session, which suggests that attaining basic competence in operating either system was not a very difficult task for these subjects. However, cognitive, perceptual, or motor impairments may dramatically affect an individual's ability to achieve basic competence, and additional research is required to assess this.

Both subject groups continued to improve across all 10 trials with each system. However, the data suggest that expertise with the Letters-only system may have developed more quickly, since one subject in each group began to plateau in their improvement with the Letters-only system, which was not observed in any Letters + WP users. This is not surprising since only the letter matrix must be mastered in the Letters-only system, while expert use of the Letters + WP system requires knowledge of the letter positions, anticipation of the word list contents, as well as development of an efficient strategy for deciding which type of item to select next. Further research that uses transcription texts with less variation and observes skill development over a longer time course would be necessary to gain a more complete understanding of learning differences.

Text Generation Rate

The differences between text generation rates with and without word prediction were not statistically significant, with the average gap between systems being about 8% (see Fig. 4). A major factor contributing to this result was the extra time required to make each selection with the Letters + WP system, which counteracted the benefit of having fewer selections to make. As with any empirical study, however, there are several methodological factors that may have had impact on these results.

Implementation of Letters + WP System

While the Letters + WP system used was not perfectly analogous to any particular commercial system, its characteristics were comparable in many ways to existing systems. The system dictionary contained a high percentage of the words to be entered, and the keystroke savings of 42% offered by the system is similar to that measured for two commercial predictive systems (8). However, it is not quite as high as the keystroke savings of 50% or more reported for some users of word prediction (10). One reason for this may have been that an adaptive prediction algorithm was not used in the Letters + WP system, so the word lists seen by subjects were fixed throughout the study. On the other hand, the fixed lists may have facilitated subjects' ability to anticipate when a word would be in the list, leading to a shorter selection time and faster learning rate. Additionally, the time required by the Letters + WP system to update the list was imperceptible. Clinical experience suggests that at least some commercial systems have a noticeable system response time, which may have a negative impact on a user's performance.

Experimental Protocol

Possible limitations in the specific protocol used include the type and number of subjects, the instruction provided to them, the time course of the study, the text blocks used, and the choice of single switch scanning as the input method. Several of these factors are discussed in greater detail below.

Able-bodied subjects were employed to provide a baseline for what can be achieved when impairments do not affect system use. Individuals who have impaired ability to use a single switch may achieve slower rates, but rates for users who have little or no motor impairment relative to hitting the switch may be similar to those found here. Additionally, these subjects were highly literate and had no cognitive or perceptual impairments, so these results should not be generalized to users who do not share these characteristics. For example, individuals who have difficulty spelling may improve both in text generation rate and quality of text when using a word prediction system (19).

Another potentially influential factor is that subjects had to decide for themselves when to use the word list, a cognitive process that may have slowed their speed with Letters + WP, particularly in the early trials. This situation is not necessarily unrealistic, however, and it also revealed some interesting differences in subjects' use of the word list. For example, one subject in Group B consciously set the timing parameters slow enough to ensure that there was enough time to find a word on its first appearance in the list, thereby maximizing his keystroke savings. Others in this group, who had slower text generation rates, appeared to be less systematic in their approach. Observations such as these raise questions for future research regarding the nature of strategies that actual users develop for themselves, the possibility that specific strategy instruction could improve performance, and the potential for identifying optimal strategies based on the system configuration and the motor, cognitive, and perceptual abilities of the intended user.

A third important limitation relates to the time course of the study. As noted above, subjects' performance with both systems was still improving when data collection ended, so it is difficult to draw a conclusion on long-term performance with word prediction relative to letter-by-letter spelling. With long-term use of word prediction, for example, some users may develop the ability to anticipate the contents of the word list accurately, thus reducing the necessary search time and improving overall performance relative to letter-by-letter spelling. The use of fixed prediction lists as opposed to adaptive ones may facilitate this learning. Further research is needed to test these hypotheses and to compare expert performance with and without word prediction.

The final methodological issues to be discussed involve the use of single switch scanning as the input method. First, because system timing parameters have a strong influence on the rate that can be achieved with a scanning system, user skill is confounded to some extent with the way in which parameters are adjusted. For example, the possibility exists that some subjects could have gone somewhat faster than they actually did in some trials, due to a conservative approach to parameter adjustment. A second aspect of single switch scanning is that selecting letter items in Letters + WP required one more switch hit than in Letters-only. However, due to the keystroke savings provided by Letters + WP, the average number of switch hits per character was actually lower than the Letters-only system. A third issue in word prediction with scanning is that the number of scan steps saved is generally lower than the keystroke savings, which reduces the benefits offered by word prediction.

This is because the relative frequency of each letter changes when both letters and words are being selected, so the frequency-based matrix is less effective. In the case of the Letters + WP system, the savings in scan steps offered by the system was 28%, with a keystroke savings of 42%. While it is difficult to estimate the impact of these three factors, it should be noted that they are general issues with any scanning word prediction system, rather than idiosyncracies specific to the experimental systems used here. Subsequent studies involving direct selection are planned, as a means of eliminating these factors.

A Cost–Benefit Model of Word Prediction

These methodological considerations do limit the ability to generalize these results to other situations. There are certainly other conditions that would yield a different result, as suggested by reports that word prediction can lead to a large improvement in text generation rate (19). To reconcile these apparent empirical contradictions, our goal is to identify underlying factors that determine when word prediction enhances rate and when it does not. As a step toward this goal, the discussion below outlines some of the principles that govern user performance.

The focus here is on user performance as measured by text generation rate, although there are additional factors that may contribute to the ultimate success of any AAC system. For example, for some users, word prediction's facilitation of spelling may outweigh considerations of sheer speed. Additionally, improving motor efficiency may reduce fatigue for some users, allowing them to work longer or more comfortably. Finally, a user may just have a personal preference for a particular system. However, even though a focus on text generation rate may be a simplification of a complex picture, it remains an important factor for many users and therefore deserves careful attention.

A useful way to consider the effect of word prediction on text generation rate is as a balance between cost and benefit. We have seen that using word prediction can decrease the rate at which items can be selected, and this can be considered to be the cost of word prediction. A primary benefit is that fewer item selections must be made. The exact amount of keystroke savings that can be achieved depends on the contents of the prediction dictionary and algorithm used, as well as the text to be generated (8,20). The relative size of cost and benefit determines the net impact of word prediction on text generation rate.

A mathematical expression for this relationship can be derived as follows (21). Cost is defined as:

$$\text{Cost} = \Delta I = \frac{I_{lo} - I_{wp}}{I_{lo}} * 100, \qquad (1)$$

where I_{wp} is the item selection rate with word prediction and I_{lo} is the item selection rate with letters-only (i.e., without word prediction). The benefit in keystroke savings is defined as:

$$\text{Benefit} = \Delta K = \frac{N_c - N_i}{N_c} * 100, \qquad (2)$$

where N_c is the total number of characters generated and N_i is the number of items selected using word prediction. The percent gain in text generation rate is defined as:

$$\text{Net gain} = \Delta R = \frac{R_{wp} - R_{lo}}{R_{lo}} * 100, \qquad (3)$$

where R_{wp} is the rate with word prediction and R_{lo} is the rate with letters-only.

To determine how cost and benefit affect text generation rate gain, text generation rate is expressed in terms of item selection rate. When word prediction is not used, the only items that can be selected are individual characters, so:

$$R_{lo} = I_{lo}. \qquad (4)$$

When word prediction is used, each item selected generates an average of N_c/N_i characters. The item selection rate is multiplied by this ratio to get the average text generation rate, so:

$$R_{wp} = \frac{I_{wp}}{(1 - \Delta K/100)}. \qquad (5)$$

Substituting Equations 4 and 5 into 3, the final cost–benefit equation is obtained:

$$\Delta R = \frac{(\Delta K - \Delta I)}{(100 - \Delta K)} * 100, \qquad \text{or}$$

$$\text{Net gain} = \frac{(\text{benefit} - \text{cost})}{(100 - \text{benefit})} * 100. \qquad (6)$$

The form of the equation reveals the simple guideline that use of word prediction will only enhance text generation rate if the benefit in keystroke savings achieved by the user exceeds the cost in item selection rate.

Applying the cost–benefit relationship to this study, it can be inferred that the cost and benefit were roughly the same, since word prediction did not significantly enhance text generation rate. Over the 10 trials of the first half, the cost in decreased selection rate (40.5%) exceeded the benefit in keystroke savings (34.8%), so the text generation rate for Letters + WP was slower than that for Letters-only. In the second half, the benefit in keystroke savings (36.4%) was a little larger than the selection rate cost (30.9%), which led to a small rate enhancement for the Letters + WP users.

There are several potential sources for the costs incurred by these subjects while using the Letters + WP system and by users of word prediction systems in general. Many of these cost factors are specific cognitive and perceptual activities required to use the system. The most commonly cited are the visual search of the list and the subsequent decision about whether the list contains the desired word. A less obvious source of cognitive load is the processing involved in planning use strategies and guiding overall activity (7,22,23). For example, a user may spend time before each selection deciding whether or not to search the list, based on the perceived likelihood that the word will be in the list. Delays may also occur when a word that the user expects to be in the list is not there, which may correspond to the time required to replace the interrupted plan of action with a new one (13). The times associated with each of these processes are generally short (on the order of a few hundred milliseconds), but taken together they can result in an appreciable cognitive time cost (7).

The use of word prediction may also increase the time involved in the motor aspects of making a selection. This may stem from additional motor requirements, such as the extra switch hit required by many predictive scanning systems. In some cases, there may also be an interaction between the cognitive requirements of word prediction and the user's motor abilities. One proposed mechanism for this is that the cognitive load may decrease resources available for motor planning (24). A second possibility is that the motor movements used to search the word list (e.g., head and neck movements) could result in motor reflexes or changes in muscle tone that then affect the difficulty of the subsequent selection movement. Further research is needed to investigate the nature and prevalence of these effects.

A third source of increase in selection time, specific to scanning systems, is the larger number of scan steps required per selection with word prediction. This means that the user must wait longer for the scanning highlight to reach the desired item, on average. In this study, the effect was fairly small; the average number of additional scan steps per selection is 0.35, which, at a scan speed of 500 milliseconds, corresponds to 175 milliseconds of extra time per selection.

These cost factors appear to have differentially affected the two subject groups in this study, since Group B, who used the Letters + WP system in the first half of the study, incurred a higher cost than Group A, who used it in the second half. The higher cost of using Letters + WP in the first half was due at least in part to the cognitive demands of using Letters + WP without any prior experience. In contrast, when Group A subjects used the Letters + WP system, they had 10 trials of prior experience using the letter matrix, which reduced the cognitive demands of making letter selections. This may have facilitated the development of more consistent and efficient strategies for using the word list, since more resources could be devoted to that aspect of the task, resulting in a higher keystroke savings and a better performance relative to Letters-only than was observed in the first half of the study.

The cost of word prediction may be reduced with practice, as suggested by the result that item selection times decreased faster across trials for Letters + WP users as compared to Letters-only users. It is also possible that word prediction may exact a lower initial cost in item selection rate for users whose impairments slow their rate of letter-by-letter spelling.[4] Equation 6 provides a way to simulate the impact of these hypothesized effects. For example, if cost were only 20% and keystroke savings were 35%, the improvement in text generation rate would be 23%. However, available information from this study and others does not tell us at what point and for whom this cost level can be achieved. Additional research involving a longer time course and a variety of users is necessary to determine the cost (i.e., the percentage change in selection rate) that may be expected under a range of conditions, analogous to the range of keystroke savings that has been established.

It should be noted that both cost and benefit are influenced by the way in which the user employs the word list. For example, a user may choose to focus on maintaining a fast selection rate (to reduce cost), but this may reduce the keystroke savings benefit as well, since less time can be spent searching the word list. Conversely, a user can decide to maximize keystroke savings by selecting each word as soon as it appears in the list, but this may increase cost since more time may be spent in at-

[4] In fact, two case reports suggest that selection rate may actually improve with the use of word prediction (11), although the source of this effect is unclear. For example, it could be due to spelling difficulties, or simply to variability in performance between rate measurements.

tending the list. Further research is needed to understand this interaction more fully.

Further developments toward modeling the quantitative relationship between keystroke savings, item selection rate, and text generation rate could yield a valuable tool for clinicians as well as designers. Such a model could simulate the effect of conditions for which specific empirical information is not available. For example, it could help determine how high the keystroke savings must be for a given cost in selection rate, in order for word prediction to improve text generation rate significantly. Similarly, for a given level of keystroke savings, the amount of acceptable cost could be determined and compared to the expected cost, as a means of estimating whether word prediction could yield a net improvement in text generation rate for a given individual.

CONCLUSIONS

Results from this study confirm that improved motor efficiency (in the form of keystroke savings) does not always lead to an improvement in overall text generation rate. In this case, the savings in necessary selections provided by the word prediction system was offset by the additional time required to make each selection. Further empirical work is needed to examine this cost–benefit balance under other sets of conditions, and the measurement of item selection rate as well as text generation rate is strongly encouraged. Finally, continued development of the theoretical principles behind word prediction and other rate enhancement techniques is necessary to provide coherence to diverse empirical results and to gain understanding of the many situations that cannot be evaluated empirically.

Acknowledgments: The authors thank Torii Patock for her assistance in performing this study. Preliminary results were presented at the 15th Annual RESNA Conference (25). This work was supported by the University of Michigan Rackham School of Graduate Studies and the National Science Foundation.

REFERENCES

1. Levine SP, Horstmann HM, Kirsch NL. Performance considerations for people with cognitive impairment in accessing assistive technologies. *J Head Trauma Rehabil* 1992; 7(3):46–58.
2. Vanderheiden GC, Kelso DP. Comparative analysis of fixed-vocabulary communication acceleration techniques. *Augment Alternative Commun* 1987;3(4):196–206.
3. Bray NW. A cognitive model for Minspeak. In: *Fifth Annual Minspeak Conference Proceedings*. Wooster, OH: Prentke Romich Co., 1990.
4. Light J, Lindsay P, Siegel L, Parnes P. The effects of message encoding techniques on recall by literate adults using AAC systems. *Augment Alternative Commun* 1990;6(3):184–202.
5. Soede M, Foulds RA. Dilemma of prediction in communication aids. In: *Proceedings of 9th Annual RESNA Conference*. Washington, DC: RESNA, 1986:357–9.
6. Horstmann HM, Levine SP. Modeling of user performance with computer access and augmentative communication systems for handicapped people. *Augment Alternative Commun* 1990;6(4):231–41.
7. Horstmann HM, Levine SP. The effectiveness of word prediction. In: *Proceedings of 14th Annual RESNA Conference*. Washington, DC: RESNA, 1991:100–2.
8. Higginbotham DJ. Evaluation of keystroke savings across five assistive communication technologies. *Augment Alternative Commun* 1992;8(4):258–72.
9. Scull J, Hill L. A computerized communication message preparation program that "learns" the user's vocabulary. *Augment Alternative Commun* 1988;4(1):40–4.
10. Newell AF, Booth L, Beattie W. Predictive text entry with PAL and children with learning difficulties. *Br J Educ Technol* 1991;22(1):23–40.
11. Newell AF, Arnott JL, Waller A. On the validity of user-modeling in AAC: comments on Horstmann and Levine (1990). *Augment Alternative Commun* 1992;8:89–92.
12. Gibler CD, Childress DS. Language anticipation with a computer-based scanning aid. In: *Proc. IEEE Computer Society Workshop on Computing to Aid the Handicapped*. New York: IEEE, 1982:11–6.
13. Treviranus J, Norris L. Predictive programs: writing tools for severely physically disabled students. In: *Proceedings of 10th Annual RESNA Conference*. Washington, DC: RESNA, 1987:130–2.
14. Caves KM, Gross K, Henderson K, Minson AA. The use of integrated controls for mobility, communication and computer access. In: *Proceedings of 14th Annual RESNA Conference*. Washington, DC: RESNA, 1991:166–7.
15. Witten IH, Cleary JG, Darragh JJ, Hill DR. Reducing keystroke counts with a predictive computer interface. In: *Proc. IEEE Computer Society Workshop on Computing to Aid the Handicapped*. New York: IEEE, 1982:3–11.
16. Rosen MJ, Goodenough-Trepagnier C. Factors affecting communication rate in non-vocal communication systems. In: *Proceedings of 4th Annual RESNA Conference*. Washington, DC: RESNA, 1981:194–5.
17. Dabbagh HH, Damper RI. Average selection length and time as predictors of communication rate. In: *Proceedings of 8th Annual RESNA Conference*. Washington, DC: RESNA, 1985: 404–6.
18. Lessenberry DD. *College typewriting*. Cincinnati: South-western Publishing Co., 1975.
19. Newell AF, Arnott JL, Booth L, Beattie W, Brophy B, Ricketts IW. Effect of the "PAL" word prediction system on the quality and quantity of text generation. *Augment Alternative Commun* 1992;8(4):304–11.
20. Swiffin AL, Arnott JL, Newell AF. Adaptive and predictive techniques in a communication prosthesis. *Augment Alternative Commun* 1987;3(4):181–91.
21. Koester HH, Levine SP. A model of performance cost versus benefit for augmentative communication systems. In: *Proceedings of 15th Annual Conference IEEE Engineering in Medicine and Biology Society*. New York: IEEE, 1993, 1303–1304.

22. Card S, Moran T, Newell A. *The psychology of human-computer interaction.* Hillsdale, NJ: Erlbaum Associates, 1983.
23. Olson JS, Nilsen E. Analysis of the cognition involved in spreadsheet software interaction. *Human-Computer Interaction* 1988;3(4):309–50.
24. Soede M, van Balkom HLM, Deroost G, van Knippenberg FHM, Kamphuis H. New possibilities for enhanced keyboard input for the handicapped. *J Med Eng Technol* 1989; 13(1/2):34–6.
25. Horstmann HM, Levine SP. Learning and performance in scanning systems with and without word prediction—report on a pilot study. In: *Proceedings of 15th Annual RESNA Conference.* Washington, DC: RESNA, 1992:299–301.