Evaluation of "Predict": An investigation and a follow-up study of a Swedish word prediction program

Magnuson, T.

journal: STL-QPSR
volume: 35
number: 4
year: 1994
pages: 001-020

http://www.speech.kth.se/qpsr
Evaluation of "Predict": An investigation and a follow-up study of a Swedish word prediction program

Tina Magnusson

Abstract
The objective of this study has been to investigate a word prediction program (Predict) as an aid for individuals with writing disabilities. Time saving and effort saving are chosen as parameters to indicate improvement and the relationship between the two parameters is discussed. In a second part of the study the potential to use Predict as a support for spelling and sentence construction is presented. Seven persons with motor disabilities participated in the study. Two of the subjects also evidenced writing difficulties of a dyslexic type. The result is that the degree of improvement relating to speed and efficiency varied considerably among subjects, depending on the underlying writing abilities and strategies. Evidence for Predict as a powerful tool for improved spelling and sentence construction is presented.

Introduction
The inability to communicate effectively through writing and reading influences a person's total development and quality of life. There may be many different reasons for writing and reading problems, such as physical impairment, spelling difficulties or poor linguistic skills. The field of augmentative and alternative communication has developed during the two last decades to support persons with speech and/or writing disabilities and to offer alternative means of communication such as signs, symbol systems and electronic communication devices.

The development of technical aids has progressed rapidly but insufficient technology in itself has not been the only obstacle to efficient communication. Technical research and development needs to be accompanied by the development of scientific methods for learning new communicative strategies, training procedures and evaluation of the devices, measurement of the effects and interpretation of the results. The importance of investigating the real effect of an aid cannot be overestimated and the need for information that comes from this type of research is of great value. (Batavia & Hammer, 1990).

Word prediction is one method of making the writing process easier and less effortful. There are a number of different word prediction programs, mostly in English, for which the basic function is that the program "guesses" what the user might want to type next according to what is previously written and gives a list of possible suggestions. Word prediction has initially been developed and evaluated as an aid to reduce the number of keystrokes and, thus, save time and effort (for a review, see Higginbotham et al., 1992; Higginbotham, 1992; Millar & Nisbet, 1993). The relation between the number of keystrokes and time saving has also been investigated. It was
found that a decrease in the number of keystrokes may not automatically mean time saving (Horstmann & Levine, 1990; 1994).

A second effect of word prediction programs is that some individuals with writing disabilities of a dyslexic type get help with spelling, word order, inflections etc. One can assume that this is because it is easier to choose between correctly spelled words than to produce them. For this reason, word prediction programs have been used and evaluated as spelling support for individuals with reading and writing problems (Newell et al., 1992).

A Swedish word prediction program (Predict) was developed (Hunnicutt, 1986) and was later rewritten and combined with a word processor. In a pilot evaluation study with four subjects (Magnuson & Hunnicutt, 1990) we compared Predict to the subjects' ordinary way of writing, and measured time saving and efficiency. We found no significant time saving, but reduction in the number of keystrokes in two subjects. It is possible, though, that the reason for the lack of improvement was due to several small programming errors, to the fact that the Subject Lexicon could not be corrected before updating and to the selection of the subjects. Three of the subjects were probably too skilled and trained on their ordinary writing systems to benefit from the program. Like Horstmann & Levine (1990), we found that a reduced number of keystrokes does not automatically mean time saving since the writing strategy and process is changed. The procedure when using a prediction system is generally to type, look at the screen, make a decision, look back at the keyboard and select the desired word. This procedure can actually increase the necessary time spent. Some individuals will save both keystrokes and time whereas other individuals save only keystrokes. Keystroke saving in itself, however, may imply less physical effort. The support the program can give was shown to be very dependent on the individual. At this point, then, no general criteria could be given for prescribing the program.

When an upgraded version of the program became available, we decided to evaluate it with modified training and testing procedures. We were interested in measuring time saving, which was calculated as the number of output characters produced during a given time, and efficiency, which means a decrease in the number of keystrokes for a given text. In this study, two subjects out of eight happened to have writing problems due to both physical impairments and reading and writing difficulties. During testing we found that efficiency and time saving were not the only benefits gained from the program but that aspects such as better spelling and sentence construction were also present. We found that other research groups noticed an effect of spelling assistance by similar programs used with school children (Newell et al., 1992). As we had all the written material collected for our efficiency study, we decided also to evaluate the linguistic skills in the written performance of the two subjects with linguistic problems. We also looked at spelling and word choice assistance as well as qualitative aspects such as intelligibility and general style.

In summary, the questions we asked in the present study were as follows: Is the number of keystrokes (effort) reduced with Predict? To what extent is the number of keystrokes related to time? Can word prediction support spelling and sentence construction and improve the overall impression of the written output?
Method

Description of the program

The word prediction program Predict, which has been under development since 1983, was originally intended to be an aid to enhance the speaking and writing rate of severely physically impaired individuals. It was rewritten as a predicting word processor in 1989 (Hunnicutt, 1989) and is run on IBM compatible PCs. It suggests (predicts) one to nine likely candidate words from the beginning letter(s) of a word. The words are shown in a box right under the entered letters and it can, thus, be possible to select a complete word with just one or two keystrokes. The candidate words are chosen either from a variable-sized frequency ranked lexicon (Main Lexicon) or from a list of lower frequency words previously used (Subject Lexicon), thus taking both frequency and recency into account. There is also a Word Pair Lexicon that, for some common word sequences, will suggest the words most likely to follow another word, thus saving keystrokes for both the second word and the intervening space. A keystroke is also saved by automatic engagement of the Shift Key after punctuation. The lexicon can be updated when desired, so that it becomes individualised. As soon as a space or punctuation mark is entered after a word, this word is added to the Subject Lexicon. The Subject Lexicon can easily be edited and corrected which is an important feature for individuals with spelling difficulties who otherwise would get an increasing number of misspelled word predicted.

Design

As individuals with disabilities do not form a homogenous group and the very disabilities are diverse, a group study with matched controls is not a suitable design (McReynolds & Kearns, 1982). This study was accordingly carried out with a single subject experimental design. Written texts were only compared within a person's own production. A baseline was established with the subjects' ordinary writing system; texts were then produced with Predict immediately after introduction of the program and after two months of use. A final test after another two-month period was administered.

Subjects

To recruit subjects for the study, a letter was sent to speech pathologists and other professionals who worked with individuals with congenital impairment in the Stockholm and Uppsala regions. The criterion for being a subject was that the person's writing disabilities were due to motorical impairment. In the enquiry we mentioned that the program could provide some spelling assistance but it was not a requirement that the subject had difficulties with spelling or sentence construction. Computer experience was not necessary. The enquiries to speech pathologists resulted in eight subjects, three male and five female, testing the system. The subjects had moderate to severe motor impairment, six with cerebral palsy and one with a muscular disease. Two of the subjects also happened to have writing problems due to difficulties with spelling and sentence constructions.
Procedure

According to a single-subject experimental design the study consisted of a baseline followed by three tests with Predict; an A-B₁-B₂-B₃ design where A indicates a baseline test with the subject's ordinary writing system, and B₁, B₂ and B₃ indicate the first, second and third tests with the Predict system.

Baseline test (evaluation of the subject's pre-test writing abilities)

The baseline test consisted of two tasks a) to copy from a given text, (Test A: copy), and b) to write about a freely chosen topic, (Test A: free). The title of the free text was to be decided before the test started to prevent the writer from spending too much of the test time thinking about the topic. Time was used as a constant and was set at 30 minutes for each of the two texts the subject had to write at this session. The time could be altered according to the subject's wishes, but was to be kept constant within the same subject. The number of characters in text the subject produced was counted to see how far in the text they came. This test was done to measure and evaluate the subjects' pre-test way of writing and was intended to be a point of reference for later tests. The subjects' pre-test writing situation, their attitudes and wishes were assessed through an interview.

Introduction of the system

The system was introduced at the same session as the baseline test was administered or at a separate session before the start of the training period. Some time was allowed from the introduction of the program to the actual start of the first training period in order to become familiar with the system and to have the opportunity to ask questions. The time for introduction was set to two weeks.

The period of usage

The use periods were to be long enough to allow the subject to get used to the system. It was necessary to have a certain amount of training to be able to shift easily between looking at the keyboard and the screen, to make rapid decisions or, when it is more efficient, to type a full word rather than accepting a prediction, to learn details regarding periods, commas and other punctuation, to handle the different lexica and files in an optimal way, and to give enough time for the lexica to become somewhat individualised. The use period was set to four months with a test after two months.

The tests

The tests were administered at three separate sessions. Before the period of usage started there was a test session with two tasks (Test B₁: copy and Test B₁: free) with Predict to determine the subject's performance with no training at all. The second test (Test B₂: copy and Test B₂: free) was administered after two months of training and the third one (Test B₃: copy and Test B₃: free) after four months of training. The time allowed for each test was long enough to allow for resting periods (which were measured) and external disturbances. This was intended to reflect the normal writing situation. The time set for each test task was 30 minutes. This time could be altered,
but was always the same as at the baseline tests for each subject. The tasks were also the same as at the baseline test, i.e., one text to copy and one text about a freely chosen topic. The copy texts were different at the different sessions but were taken from the same chapter of the same book.

**Evaluation**

The tests were to be evaluated regarding both time saving and efficiency but also as regards differences in the linguistic quality of the written output.

**Results**

**Subjects**

All subjects except subjects B and D continued using Predict until the end of the study, but with varying degrees of fulfillment. Unfortunately, subject illness and failure in copying some log files due to computer errors made the study not as complete as planned, but as every subject is compared only to him/herself it has been possible to draw some conclusions from each person's contribution. A review of the results, taken subject by subject, follows. The results are analysed using within-subject comparisons. The presentation of the subjects follows no particular order.

**Efficiency and time saving**

In order to get information about the writing process, a log file was analysed. This log file registration provided information about the time and type of each keystroke. The registrations were counted manually. There were several parameters that were measured and compared. To measure time saving we calculated how far in a text a subject came during a given time, i.e., how many characters were entered which appeared in the final written output. To evaluate the efficiency we measured the number of uncorrected errors, extra keystrokes due to capital letters, the number of corrections with backspace or arrows, the number of corrected errors, the total number of keystrokes excluding corrected errors and corrections, the total number of keystrokes including corrected errors and corrections and number of accepted predictions. These parameters were compared to the total number of characters in text. Thus, we were able to measure the relation between the amount of entered text and the number of keystrokes needed. In the figures, "characters in text" means the final written product; "keystrokes" means the total number of keystrokes including extra keystrokes due to capital letters, corrected errors and corrections. "Accepted predictions" means the number of times the subject has selected a predicted word. All texts were normalised to 30 minutes of writing. Missing data due to failure in copying the log files are marked n.a. (not available) in the figures.

**Results from the efficiency and time saving evaluation**

**Subject A**

Subject A was a male who usually writes on an Apple Macintosh. He did not really like the idea of working on a PC compatible but was eager to be instrumental in
product development. He liked to try out new programs and was looking forward to a Mac version of Predict. He was a fast, experienced writer and not very motivated to learn a new typing strategy. The B1:copy test results were unfortunately lost due to system error during transferring the files to diskette.

![Figure 1. Subject A, copied text.](image)

![Figure 2. Subject A, free text.](image)

As can be seen in Figure 1, the number of keystrokes needed always exceeded the number of entered characters, i.e., the final written output. He hardly used the predictions at all. Subject A did not improve with Predict, neither in time nor efficiency. He did not produce more text nor did he use less keystrokes. In the free text (Figure 2), the results improved after two months use although he did not take advantage of the predictions. This may indicate that learning a new program takes time in itself regardless of what the program does. There was only one occasion when the number of keystrokes did not exceed the number of characters in text, i.e., he saved keystrokes by using the predictions only on one test (Test B1:free). This was the occasion when he was least productive, which indicates that this was a more laborious method for him, requiring more of his concentration.

**Conclusion:** Subject A types somewhat faster both at copying and at free writing with his ordinary program. He did not use the predicted words much and it was probably faster for him to type out the words. He was too fast a typist to benefit from the program and he did not need the spelling help. He liked the idea of the program, however, and thought it could be of great value for other users.
Subject B

Subject B was not one of the initial subjects, but asked if she could participate. She wrote only the baseline test (A1:free) and the first test with predictions (B1:free) so she did not follow the complete study design. It can more be regarded as an example of her writing with and without predictions. She had no spelling problems but rather a kinesthetic problem. When her hands change position she does not feel it and therefore sometimes gets everything wrong when she uses the touch method of typing. This problem never occurred when she used Predict. To be forced to look at the keyboard and the screen helped her to keep her hands in the right position. When she wrote the test without Predict she made forty errors, but only one error when writing with Predict. An additional gain was the decrease in the number of keystrokes, as can be seen in Figure 3.

Figure 3. Subject B, free text

Conclusion: Although Subject B is not extremely slow and has no spelling difficulties, she felt the program helped her rather much because it forced her to use a more efficient typing strategy and she was able to type the same amount of text with fewer keystrokes.

Subject C

Subject C was an experienced writer. She was highly motivated to learn the program and tried to follow the instructions to use the predictions as often as possible. This may have had much to do with her wish to be a helpful subject as well as her assumption that the program would be efficient.

Figure 4. Subject C, copied text.
As shown in Figure 4, the copying test indicates keystroke saving related to the amount of written text, but the total production of text decreases with Predict. This means a gain in efficiency but not in time saving. The log file for test B3 was lost due to computer error when transferring the file to diskette. She reported that similar incidents often happened. The test of free text (Figure 5) shows that text production reached the same result after the entire training period as it achieved in the baseline test. In the baseline test, the subject wrote about a topic that she knew very well and very often wrote about. Many of the sentences were probably already mentally "preprogrammed". The result was extraordinarily good according to the subject herself, and the baseline test (Text A: free) may therefore not be representative. The total number of characters in text may have been fewer at the baseline test with a different selection of topic. The number of characters in text is about the same in B3: free but with far fewer keystrokes. Since all baseline tests (test A) were written with the subjects' ordinary writing systems there were no log files. The information about deletions and extra keystrokes had to be collected by ear and hand, looking over the subjects' shoulder which could sometimes feel awkward. That is the reason why there are no data on the total number of keystrokes from the baseline test (this was also what happened with subject H). The total number of keystrokes cannot be less than the number of characters in text (588). Comparing with her other writing, one can estimate 50 extra keystrokes, which will give 640 compared to less than 400 keystrokes for test B3: free. With training and adaptation of the lexica, speed and number of accepted predictions goes up noticeably and the total number of keystrokes goes down moderately. The total number of keystrokes has not changed very much, which indicates that accepted predictions have been substituted for manually entered letters.

**Conclusion:** The number of characters in text per minute increased and the total number of keystrokes decreased. With more training and further adaptation of the lexica, she would probably benefit even more from the program. The time saving may not be as important as the gain in efficiency indicating less effort.

**Subject D**

Subject D was a male who, due to a muscle disease, was very slow and very easily fatigued. As he could write only for about ten minutes at a time, he preferred to do only the free text and suspend the copying task. During this last test, he appeared not to
be at all fatigued after 10 minutes and was allowed to continue past the time set by the baseline test to 30 minutes. Therefore, the total number of characters in text (336) typed in 30 minutes had to be divided by three (112 characters in text) in the analysis to give an equivalent to the ten minute baseline test.

As indicated in Figure 6, there was no gain in efficiency and in test B2:free the number of keystrokes exceeded the number of characters in text. One influence was that he experimented with the words. Of the total number of keystrokes, 28 were due to accidental engagement of the repetition function of the computer. The difference in real numbers may not seem large, but by comparing the three texts the improvement is salient. It was not until the last test that the subject began to use the advantageous features of the program after which his results increased. There had been very little training because of subject illness but the writer himself thought the program was good and is interested in continuing training in the future.

**Conclusion:** Subject D is extremely slow and very easily exhausted so it is to be assumed that this program is going to help him significantly.

**Subject E**
Subject E was a female with a very severe motor disability. She typed with a headstick and sometimes used a separate scanning keyboard (Thot) with knee switches.
Figure 8. Subject E, free text.

The copying test (Figure 7) shows that subject E's results (number of characters in text) never went above her baseline results. Acceptance of predicted words slightly reduced the number of keystrokes compared to the amount of text. As can be seen in Figure 8, the free text showed a slightly different result from the copying test. She took more advantage of the predictions. The amount of entered text always exceeded the total number of keystrokes which indicates that there was less effort with the program. On one occasion she wrote more text with the Predict program which means that it also was time saving.

Conclusion: Subject E has very severe motor disability and types with a mouthstick so we expected her to benefit from the program. However, there was no time saving except for Test B2:free, where the number of characters in text was larger than both the number of keystrokes for the same text and the number of characters in text without predictions. After the end of the project, E preferred to go back to her ordinary word processor.

Subject F

Subject was a female with very slow typing. Her motor disabilities did not account for all her writing difficulties and her slowness puzzled the staff and professionals around her. It was never made clear whether she liked the program or not but she fulfilled the task very cautiously.

Figure 9. Subject F, copied text
Figures 9 and 10 show that for all types of tests the total number of keystrokes always exceeded the amount of entered text. Either she never learnt to use the predictions or did not have any use for them.

Conclusion: F was extremely slow although her motor abilities would allow her to be faster. Her writing process would probably be slow regardless of the writing system she used. The only effect from the new system that did not work in the way she was accustomed was actually to slow her down.

Subject G

Subject G was a male who was a rather fast writer in spite of his motor disabilities. He learnt the principles of the program from the beginning but often preferred to type out words he knew how to spell. During the study he moved to the southern part of Sweden and the tests were thereafter conducted by the speech therapist at the communication center where he studied.
In Figure 11 we can see that there is a slight time saving and savings in efficiency in the B1:copy and B2:copy tests. In the Test B3:copy the time saving was more noticeable (23%), but as the log file could not be retrieved we do not know if this savings was due to the number of accepted predictions. G's feelings about the program were very positive. He thought that it gave him more confidence in his own writing and that he now had the courage to let other people read his texts. This has to be compared to the measurable results of the tests, however, which did not show an overwhelming improvement, as noticed in Figure 12. The amount of entered text was about the same in the different tests of free text, but is somewhat higher with Predict than without (from -2% to 12%). The total number of keystrokes is also about the same (from 3 to 13% more). The ratio between entered text and number of keystrokes is largest in the baseline test and in test B3:free, where he does not use the predictions very much, which indicates that he used fewer keystrokes per entered character of text using the predictions.

Conclusion: Although the difference in result is not very large, there is enough improvement to indicate the advantages with Predict. A second gain is G's very positive attitude towards Predict and the fact that he now writes much more.

Subject H

Subject H was a female who showed no linguistic difficulties when speaking but evidenced a dysarthria related to her motor impairment. She had writing difficulties due to both motor impairment and difficulties with spelling and sentence construction. She was very motivated to learn and to use the prediction program and had earlier been looking for something that could help her with her writing difficulties.
In the copying tests in Figure 13 we can see that she performed noticeably better with Predict, both in speed and efficiency. There was a 27, 35 and 88% time saving in Test B1, B2 and B3, respectively. The ratio between characters in text and keystrokes was 1:1.21; 1:0.84; 1:0.86; 1:0.80 which means that the keystroke saving also increased with training. For the same reasons as for subject C, there are no data on the number of keystrokes in the baseline test (test A). In the free writing, we can see in Figure 14 an improvement in speed in the last test. The number of accepted predictions follows the general pattern, i.e., "good" sessions, with a high degree of speed and efficiency, often contain a high degree of accepted predictions.

**Spelling ability and reader judgement of the texts**

In the analysis that aimed at evaluating Predict as a linguistic support for writing, three subjects were included: Subjects G and H who had difficulties with spelling and text construction and subject E, who was included as a control. As free writing was the task that best reflected a natural writing situation, only the free texts were analysed for this purpose. Figures from the spelling analysis protocol will be presented as well as results from the reader judgement.

To measure spelling, eight parameters were used: omitted letters, extra letters, reversed letters, substituted letters, omitted word boundaries, extra word boundaries, upper/lower case mistakes and incorrect inflections. (Since we have not yet analysed the texts morphologically, incorrect inflections were included in the spelling errors.)

The number of spelling errors was counted for each text.

To evaluate intelligibility and the general style of the text, ten readers that were naive to the purpose of the study assessed and compared the texts. Baseline and test texts from the three subjects were arranged in a random order and given to the readers who could read the texts any number of times. Each text was followed by the four questions seen below. The readers were then asked to judge the texts by rating these four factors on a scale from one to ten.

1. How easy is it to **understand the text**? (1 = totally unintelligible; 10 = totally intelligible)
2. What is your opinion of the **grammar** of the text? (1 = very poor grammar; 10 = grammatically correct)
3. What is your opinion of the author's ability to "dress his thoughts in words"? (1 = very poor vocabulary; 10 = very abundant vocabulary)
4. What is your **general impression** of the text? (1 = badly written; 10 = very well written)

**Results from spelling error analysis and reader judgement**

There are no reader judgement data for test B3:free as the reader judgement was conducted before the end of the study, i.e., after test B2. In the text, spelling error score is given with test B3:free included.
**Subject E**

Subject E had no linguistic difficulties but was included in the reader judgement as a comparison. She made no errors in the A1:free, one error in B1:free, no errors in B2:free and three errors in B3:free. Although she rated very high on intelligibility and grammar, the overall impression was lower. The scores shown in Figure 15 are the results from the four questions relating to reader judgement on the left and the means from the four factors on the right.

![Figure 15. Subject E, differentiated result and mean from reader judgement](image)

As confirmation of the reliability of the testing method it was noted that the lowest result in efficiency and time saving (Figure 8) is to be found in the same test as the test that gets the lowest result in the reader judgement (Figure 15) and that the test where she was most productive also gives the best score on the reader judgement (B2:free where both production and the score for general impression is high). She seemed to be very sensitive to external disturbances which may have caused the differences.

**Subject G**

Subject G had severe difficulties with both spelling and sentence construction. He was very insecure about his own writing and did not want to use it for functional purposes. His copying abilities were fairly good (eight errors for A1:copy; eighteen for B1:copy; twelve for B2:copy and five for B3:copy), but in the case of free writing he had serious difficulties. He worked relatively fast and rarely went back to look for errors or to correct errors. He made 54 (12%); 45 (9%); 48 (11%) and 32 (7%) errors in the four free texts, i.e., he made fewer spelling errors with the help of predictions.

![Figure 16. Subject G, differentiated result and mean from reader judgement](image)

The means from the four aspects provided by the results from the reader judgement are shown in Figure 16.
Intelligibility clearly improves with Predict. Grammar and general impression also improve but the mean for vocabulary goes down. The reason for this decrease in vocabulary score may be related to his tendency to make an incorrect word selection. Thus the words were correctly spelled but inadequate. The right side of Figure 20 shows the mean from the four factors added together. A high number of spelling errors is related to a low score on the reader judgement, as expected. The spelling errors were fewer with Predict and the score from the reader judgement was also higher. This shows that the biggest help for this subject was with the linguistic aspect, that is spelling. The subject himself thinks he is much more confident now in writing and that the program gives him courage to do so. There was an interesting lack of correlation of G's very positive reaction to the program and the only moderate improvement in the results.

**Subject H**

Subject H very soon discovered that Predict served as a spelling support. She was aware of her spelling mistakes and often asked how to spell certain words. She improved from 27 errors (13% of the total text) in A:free to nine errors (4%) in B1:free and ten errors (6%) in B2:free and finally to two errors (1%) in B3:free. She was very happy about this result herself and began the project of writing down her life story.

The result from the reader judgement reflected the result from the written texts very well. Intelligibility improved from six to ten on B1:free (Figure 17), which meant total intelligibility. The total result from the reader judgement for subject H can be seen on the right.

![Figure 17. Subject H, differentiated result and mean from reader judgement](image)

**Summary of results**

<table>
<thead>
<tr>
<th>Subj A</th>
<th>Subj B</th>
</tr>
</thead>
<tbody>
<tr>
<td>- male</td>
<td>- female</td>
</tr>
<tr>
<td>- cerebral palsy</td>
<td>- poor kineshtetic feedback</td>
</tr>
<tr>
<td>- relatively fast, experienced Mac user</td>
<td>- fast</td>
</tr>
<tr>
<td>- not very motivated to learn the program</td>
<td>- benefits from need to look at</td>
</tr>
<tr>
<td>- result becomes worse with Predict</td>
<td>- keyboard and screen</td>
</tr>
</tbody>
</table>
### Discussion and conclusions

#### Efficiency and time saving

The initial main function of Predict was to save keystrokes (effort). In theory it may result in a 43% decrease (with five predictions) in the number of keystrokes needed, evaluated on a transcription of one person's communication via a personal communicator (4567 words) (Hunnicutt, 1986). In this study, we have seen that keystroke saving actually took place, that the subjects really needed fewer keystrokes to produce the same amount of text. However, the greater potential savings was sometimes affected by a number of factors. The number of mistakes and corrections increased remarkably for some individuals. This increase was often due to experimentation with the program features or selection of the wrong word, which meant many extra backspaces and corrections. Thus we did not observe the 43% keystroke savings. But some decrease in the number of keystrokes, nevertheless, does mean less effort which is of great importance to individuals who find writing effortful and exhausting. To these individuals, less effort may be of greater value than time saving.

It was also assumed that with fewer keystrokes needed, the writing speed would increase. As mentioned in the introduction, it has been previously reported that keystroke savings does not automatically mean time saving (Horstmann & Levine, 1990; 1994). This conclusion was confirmed in this study where we noticed time

| Subj C | - female  
|--------|----------
|        | - cerebral palsy  
|        | - relatively fast, experienced  
|        | - motivated to learn the program  
|        | - number of keystrokes per minute increases  
| Subj D | - male  
|        | - very slow, easily fatigued  
|        | - benefits in amount of text  
| Subj E | - female  
|        | - severe motor impairment  
|        | - uses headstick and scanner  
|        | - benefits less than expected  
| Subj F | - female  
|        | - very slow in all movements  
|        | - not motivated  
|        | - does not use the predictions  
|        | - result slower with Predict  
| Subj G | - male  
|        | - cerebral palsy  
|        | - moderate speed  
|        | - reading and writing difficulties  
|        | - small gain in number of keystrokes and amount of text  
|        | - substantial benefit in intelligibility  
|        | - very positive feelings  
| Subj H | - female  
|        | - cerebral palsy  
|        | - slow  
|        | - reading and writing difficulties  
|        | - gain in number of keystrokes and amount of text  
|        | - substantial benefit in intelligibility  
|        | - improves in all aspects  

---

STL-QPSR 4/1994
saving in some individuals whereas other persons were actually slowed down. One important factor is that a changed writing strategy may consume this potential time saving, at least in the beginning. Unless the touch method of typing is used, it is necessary to look at the keyboard, look at the screen, read the predictions, make a decision, go back to the keyboard and press a selection key. Some individuals go through this process faster than they type all necessary characters whereas others type faster, depending on the basic writing strategy and writing speed. However, four subjects produced more text with Predict than without in the same amount of time. Two subjects reported that it was easier and less effortful to write with the program but that it took more time.

We expected to find that persons usually do not save all the keystrokes that are theoretically possible and that a gain in the reduced number of keystrokes automatically means time saving. What we did not expect was how difficult it would be to pre-determine who would benefit and who would not. The persons we believed would meet with our criteria for being a successful Predict user may not have improved as much as we assumed whereas someone with less severe motor impairment might make substantial gains. There may be "invisible" differences among persons, such as flexibility in changing strategy, ability to quickly move eye gaze, differences in personal styles, such as taking the time to read all predictions or choosing a prediction as soon as it is seen. Another puzzling fact is that we do not find the large improvements that we hear about from users and speech pathologists. One reason for this finding may again be the selection of subjects. One conclusion we draw from this study is that the program should not be prescribed without testing and that the prerequisites to succeed are very different for each person. At the same time our subjects were recruited on very general and unspecified criteria. The number of participating subjects is also very small due to the fact that this is a very time consuming type of study. This means that we do not get very many samples from persons with different types of underlying difficulties and abilities. One idea that has arisen from this finding is to run a similar study with the persons that have reported substantial benefits from the program to investigate their writing strategies, basic difficulties and abilities.

The subjects do not always use the program in an optimal way. They type words that are suggested and do not use the predictions. This may be due to the writer's knowledge about when it is more efficient to type the word than going through the prediction process, but it may also be a result of habit, not enough training, not understanding the potential benefits of the program, etc.

For our subjects, the program has, in general, been more efficient than it has been time saving. Most typically, only very slow typists save time, and it has been concluded that several of the subjects in this study have been too fast to benefit from the prediction program. We have been able to refine our criteria, but there is more work to be done on this issue. The program cannot be recommended without testing to persons who are slow writers. There must be information on how slow they are, why they are slow and their basic writing strategy, abilities and difficulties. Earlier investigations have indicated that prediction techniques do not contribute to more efficient writing. These studies have often been conducted with normal subjects in a
laboratory setting. This study shows that individuals with writing difficulties related to motor impairment do benefit in both effort and time. It is not the responsibility of the investigator to decide whether this seemingly small gain is significant or substantial enough for the user. With an even more careful selection of subjects, it is assumed that the result will be more convincing regarding the effectiveness of prediction techniques. The need is highly personal and individual.

**Spelling and sentence construction**

Of the two subjects who had difficulties with spelling and text construction, one showed substantial improvement with Predict which she really was able to appreciate. The other subject's improvement was not as large but psychologically it made a big difference to him.

One feature of the current prediction program that we want to use in the next study is the possibility to edit newly entered words. One problem has been that persons with reading and writing difficulties were not able to update the lexicon without help. If they did, they ran the risk of constructing a lexicon with an expanding number of misspelled words.

Linguistic support was a secondary aim of the present study and occurred without specific planning. The material is not extensive (two subjects) but the result indicates the power of prediction techniques as linguistic support for writing. These results have inspired us to conduct further studies to explore the various possibilities to improve text construction. In the present study, spelling improvement and impact on the reader has been investigated. We plan to complement this investigation with a morphological, syntactical and sentence complexity analysis.

**Acknowledgement**

The participation of the subjects, "readers" and reviews of the manuscript by Sheri Hunnicutt and Kerstin Nauclér is gratefully acknowledged. This project has been supported by the Swedish National Board for Industrial and Technical Development (NUTEK).

**References**


