1

MICRO

**National Library
of Canada**

**Bibliothèque nationale
du Canada**

Canadian Theses Service

Service des thèses canadiennes

**Ottawa, Canada
K1A 0N4**

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

**A Comparison of Comprehension Scores Obtained by
Children on Paper and Computer Versions of the Stanford
Diagnostic Reading Test**

**Monica D. Dykeman
Department of Psychology
Carleton University**

**Dissertation presented to the Faculty of
Graduate Studies in partial fulfillment of the
requirements of the Degree of Doctor of
Philosophy in Psychology.**

December, 1988

(c) Monica D. Dykeman, 1988

National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-315-51115-X

The undersigned hereby recommend to
the Faculty of Graduate Studies and Research
acceptance of the thesis,

"A COMPARISON OF COMPREHENSION SCORES OBTAINED BY
CHILDREN ON PAPER AND COMPUTER VERSIONS OF THE
STANFORD DIAGNOSTIC READING TEST"

submitted by

MONICA DYKEMAN, B.Ed., M.Ed.

in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

Chair, Department of

Thesis Supervisor

External Examiner

Carleton University

Date*

*To identify the date
of final revisions

ABSTRACT

There has been no previous research examining the comprehension of children reading from computer compared to paper. In this study, children assessed by their teachers as above-average ($n=24$) average ($n=24$) and below-average readers ($n=24$) completed the comprehension and silent reading speed subtest of the Stanford Diagnostic Reading Test (SDRT), once with pencil and paper and once on a Macintosh 11 computer with a high resolution monitor. Silent reading speed did not differ between media, however, scores on the computer version of the test declined 10% for average subjects ($p \leq 0.001$) and increased 13% for below average subjects ($p \leq 0.001$). Scores for above average subjects did not differ between media. Providing a keyboard to answer questions presented on paper did not influence the distribution of scores in a control group of subjects. A test of selective attention in which the number of spelling errors missed was 50% higher on computer for average subjects, and the number of words scanned by the below average subjects was 25% lower on computer, suggested that the computer monitor may require a greater degree of attention when reading for comprehension. The changes observed demonstrate that the medium on which reading material is presented interacts with reading level to alter comprehension scores.

The interaction between reading level and reading medium was manipulated by decreasing the difficulty of test material for grade five subjects and increasing the difficulty of material for grade four subjects. Average ($n=7$) and below average ($n=6$) grade five subjects,

completed the grade four version of the SDRT, and above average (n=24) and average (n=23) grade four subjects completed the silent reading fluency subtest of the grade five version of the SDRT. For the grade five subjects, 'easy' test material resulted in an 8% decrease in comprehension from the computer for below average readers and no change for average readers. For grade four subjects, 'difficult' material resulted in a 10% decrease in scores on computer for above average readers and a 10% increase in scores on computer for average readers. Reading level interacted with the difficulty level of the reading material on the computer indicating that reading ability must be considered when presenting curriculum material on computer. The results emphasize that caution must be used in transferring psychometric tests which require reading skills to the computer.

TABLE OF CONTENTS

CHAPTER 1	1
INTRODUCTION.....	1
Models of the Reading Process.....	3
Visual Skills in Reading	7
Visual Skills and Vision Research.....	14
Reading from Computers and Related Research.....	25
Studies with Adults.....	27
Studies with Children.....	36
The Computer and Reading Comprehension	43
CHAPTER 2	48
Description of the Present Study.....	48
STUDY 1.....	55
Method	55
Results.....	61
Summary.....	66
CHAPTER 3	68
PART 1: Reading Comprehension.....	69
Method	69
Test Environment:	75
Room Illumination:.....	75
Procedure:.....	78
Results.....	81
Order, Form, Sex and Grade Effects:	81
Reading Level Assignment:.....	83
Monitor Resolution:.....	84
Comparison Of Paper-Computer Comprehension Scores:.....	86
PART 2: Response Device	93
Method	94
Subjects:	94
Procedure:.....	94
Task Description:.....	95
Results.....	95
Summary.....	100
CHAPTER 4	102
PART 1: Reduced Reading Material Difficulty	103
Method	103
Subjects	103
Procedure:.....	103

Results.....	103
PART 2: Increased Reading Material Difficulty	108
Method	108
Subjects	108
Procedure:.....	108
Results.....	108
Summary.....	109
Study 4.....	116
Method	116
Subjects:	117
Procedure:.....	117
Task Description:.....	118
Results.....	118
Summary.....	121
CHAPTER 5	127
DISCUSSION	127
Experimental Design	127
Major Findings	129
Reading Level and Material Difficulty Interaction.....	134
Significance of the Findings	139
Direct Media Comparison Research	140
Selective Attention and Reading from Computer Monitors.....	143
Recommendations for Further Research	146
SUMMARY AND CONCLUSIONS	148
BIBLIOGRAPHY	151
APPENDICES	161

LIST OF TABLES

TABLE 1: Test for Differences Between Reading Levels On Oral Reading Rate, Oral Reading Accuracy and Total Comprehension Scores on Stanford Diagnostic Reading Test Questions.	62
TABLE 2: Within Reading Levels Contrast Of Paper Computer Scores on the Stanford Diagnostic Reading Test in Grade 4 and 5 Subjects.....	63
TABLE 3: Comparison of Teacher's Assignment and Predicted Assignment to Reading Level.	65
TABLE 4: Illumination Measures for Testing Rooms By School.....	76
TABLE 5: Analysis of Variance Test for Effects of Reading Level, Order, Form, Grade and Sex on The Subtests of the Stanford Diagnostic Reading Test.	82
TABLE 6:A Comparison of Teachers Assessments of Reading Skill and Test Assignment of Reading Level for Grade Five Subjects.	83
B Comparison of Teachers Assessments of Reading Skill and Test Assignment of Reading Level For Grade Four Subjects.....	84
TABLE 7: Test for Differences Between Monitor Resolutions On Literal, Inferential and Total Comprehension Subtests on the Stanford Diagnostic Reading Test.	85

TABLE 8: Test for Differences Between Reading Levels On Literal, Inferential and Total Comprehension Scores on the Paper and Computer Versions of The Stanford Diagnostic Reading Test Questions.	87
TABLE 9: Test for Differences Within Reading Levels On Literal, Inferential and Total Comprehension Scores on the Paper and Computer Versions of The Stanford Diagnostic Reading Test Questions.	88
TABLE 10: Within Reading Levels Contrast Of Paper-Computer Scores on the Context Prediction Subtest of the Stanford Diagnostic Reading Test, by Grade.	91
TABLE 11: Test for Differences Between Keyboard and Pencil as an Input Device On Literal, Inferential and Total Comprehension and Reading Rate and Context Prediction Subtests on the Stanford Diagnostic Reading Test.	96
TABLE 12: Within Reading Levels Contrast Of Paper Keyboard- Computer Scores on the Comprehension Subtest of the Stanford Diagnostic Reading Test.	98
TABLE 13: Within Reading Levels Contrast Of Paper Keyboard- Computer Scores on the Reading Rate and Context Prediction Subtests of the Stanford Diagnostic Reading Test.....	99
TABLE 14: Within Reading Levels Contrast Of Paper Keyboard- Computer Scores for Grade Five Subjects on the Grade Four Test of the Stanford Diagnostic Reading Test.....	104

TABLE 15: Test for Interactions Between Reading Levels, Material Difficulty and Media On Total Comprehension Subtest Scores on the Stanford Diagnostic Reading Test.	107
TABLE 16: Within Reading Levels Contrast Of Paper-Computer Context Prediction Scores for Grade Four Subjects on the Stanford Diagnostic Reading Test.	109
TABLE 17: Test for Differences Between Reading Levels On Omission, Commission and Error Scanning Rate Rate on the Paper Version of the Selective Attention Task.....	119
TABLE 18: Within Reading Levels Contrast Of Paper Computer Scores For the Selective Attention Test.	121
TABLE 19: Test for the Effect of Computer Familiarity On Differences Between Paper and Computer Literal, Inferential and Total Comprehension Scores on the Stanford Diagnostic Reading Test.	126

LIST OF FIGURES

Figure 1 : Design Schematic for Study 1 and 2.....	53
Figure 2 : Mean Scores Obtained on the Paper and Computer Modified Versions of the Comprehension Subtest of the Stanford Diagnostic Reading Test.....	64
Figure 3 : Sample Test Page Layout for the Computer Version of the Stanford Diagnostic Reading Test.....	71
Figure 4 : Sample Test Page Layout for the Paper Version of the Stanford Diagnostic Reading Test.....	74
Figure 5 : Luminance Measurements.....	77
Figure 6 : Mean Literal and Inferential Comprehension Scores Obtained on the Paper and Computer Modified Versions of the Comprehension Subtest of the Stanford Diagnostic Reading Test.....	90
Figure 7 : Stanford Diagnostic Reading Test Scaled Scores Obtained on the Paper and Computer Modified Versions of the Comprehension Subtest of the Stanford Diagnostic Reading Test.....	92
Figure 8 : Mean Scores Obtained by Grade 5 Subjects on the Paper and Computer Modified Versions of the Comprehension Subtest of the Grade 4 Stanford Diagnostic Reading Test.....	105
Figure 9 : Mean Context Prediction Scores Obtained by Grade 4 Subjects on the Paper and Computer Modified Versions of the Reading Speed Subtest of the Stanford Diagnostic Reading Test.....	110
Figure10: Mean Cumulative Difference in Scores Obtained on the Paper and Computer Modified Versions of the Comprehension Subtest of the Stanford Diagnostic Reading Test.....	113
Figure11: Sample Test Page Layout for the Selective Attention Test.....	120

LIST OF APPENDICES

APPENDIX 1 Percent comprehension score calculation.....	161
Means and Standard Deviations by Grade and Reading Level	162
APPENDIX 2 Specifications for Mac 11 monitors	166
APPENDIX 3 Reading rate for vertical paper	167
APPENDIX 4 Consent Form	168
APPENDIX 5 Keyboard Layout.....	169
APPENDIX 6 Relative Visual Performance calculation.....	170
APPENDIX 7 Within Reading Levels Contrast Of Paper - Computer Scores on the Reading Rate Subtest of the Stanford Diagnostic Reading Test; by Grade.....	172
APPENDIX 8 Comparison Of Paper - Computer Scores on the Comprehension Subtest of the Stanford Diagnostic Reading Test Without Regard to Reading Level.	173
APPENDIX 9 Comparison Of Paper - Computer Scores on the Silent Reading Speed Subtest of the Stanford Diagnostic Reading Test Without Regard to Reading Level.	174
APPENDIX 10 Absolute scores on paper and computer by grade and reading level.....	175
APPENDIX 11 Cumulative scores on computer and paper by grade and reading level.....	177
APPENDIX 12 Notched Box Plot of Absolute Scores by Grade.....	179
APPENDIX 13 Means and standard deviations by Test Form	181
APPENDIX 14: Within Reading Levels Contrast Of Paper- Computer Scaled Scores on the Stanford Diagnostic Reading Test.....	183

ACKNOWLEDGEMENTS

I would like to thank Apple Canada, especially Larry Smith, and the staff of Computer Connections for providing the equipment necessary to complete this study. I would also like to thank Daniel Walters of Micro-Achievement Centre for invaluable technical assistance. A special thank-you is due to the principals, teachers and students of the Carleton School board for their enthusiastic participation in this study and to Dr. E. Otoo of the School of Computer Science for allowing me access to the equipment necessary to complete this study. I cannot sufficiently express my appreciation to my co-supervisors, Dr. Mark Rea and Dr. Bruce Ferguson, for their advice and support during the last two years. Last, but not least, I would like to thank my family for their patience and understanding throughout this study.

CHAPTER1

INTRODUCTION

The personal computer is rapidly achieving the status of an alternative teaching device in Canadian classrooms. While there were virtually no computers in the classroom in 1976, it was estimated in 1986, that there were over 134,000 microcomputers in Canadian schools (Maynard, 1988). The increased use of microcomputers in the schools is based partly on the perceived potential of the computer to revolutionize teaching and partly on the novelty of the medium (Chandler, 1983). However, the implicit assumption underlying the use of microcomputers as an alternative teaching tool is that, given a sufficiently high resolution monitor, there should be no difference in comprehension between text read from a computer monitor and text read from a sheet of paper.

The assumption that a computer monitor is equivalent to a sheet of paper has been fostered by promotion of 'WYSIWYG' (what you see is what you get) monitors. Several reports indicate that 'WYSIWYG' monitors can improve reading speed and reading accuracy compared to older, low resolution monitors (Radl, 1980; Bauer & Cavonius, 1980; Gould, 1985), however, reading speed and accuracy represent only components of the reading process. When a comprehension component was included in one study of adult reading accuracy, the results suggested that presenting reading material on a computer monitor, compared to presenting similar material on paper, may interfere with comprehension (Bagnara, 1980). A later study, however, once again indicated that computer monitors

primarily alter reading speed rather than comprehension (Hepfner, et.al. 1985). Since the two studies used different approaches to measure comprehension, it is difficult to determine why the results differ.

The assumption of media equivalency for comprehension has been carried over into the research on the use of computers by children. Most of the research has been based on either the observations of classroom teachers or uncontrolled classroom experiments. The few studies that do provide any information pertaining to comprehension have proceeded on the basis that a direct comparison of computer and paper scores is unnecessary. Although the results of these studies are contradictory, it has been assumed that performance on computer-based tasks in an experimental setting can be compared to previous performance on paper. Several studies with children have used this approach. One study by Grocke, (1983) and another by Riel (1983), suggest that children with academic difficulties can benefit from a computer-based curriculum. Atkinson, (1974) and Candy and Edmonds (1982) report that the computer is less beneficial to children without academic difficulties than to those in need of academic assistance. Hawkins (1984) and Tucker (1983), however, reported that there is no evidence to support the proposition that the computer is a superior medium for providing children with practice in academic skills.

The current state of confusion in the literature related to the use of the computer as a reading medium appears to be due, in part,

to the methodology used to address separate aspects of computer use. Studies using adults have been aimed primarily at measuring speed and accuracy, resulting in a technological focus which draws heavily on the field of vision research. Studies with children have focused on the area of improving classroom-based skills, such as reading and spelling, and have focused on the field of reading research. Studies using adults have generally ignored the importance of the field of reading research while studies with children have failed to consider the importance of visual aspects in computer-based research. Past research, therefore, does not provide the information required to determine whether comprehension is affected when text is presented on computer monitors. If educators and researchers intend to substitute the computer as a convenient, automated alternative to traditional pencil and paper procedures, a study is required in which both reading and vision research methodologies are utilized to directly compare comprehension of text read from paper and from computer monitors.

Models of the Reading Process

Determining which aspects of the reading process are important in a study of reading from different media requires a brief review of the current theoretical models of reading. Since the area of interest is the effect of presenting reading material on media which primarily differ visually, auditory, language and cognitive skills will not be included in the review.

Three major models of the reading process currently appear to dominate the reading literature: the phonic-based, or 'bottom-up'

model (Ross 1976); the psycholinguistic reading-for-meaning , or 'top-down' model (Goodman, 1971); and the interactive-compensatory model which addresses the issue of skills interaction (Stanovich,1980). The 'bottom-up' model of reading assumes that the skills involved in reading operate in a hierarchical manner with visual information entering by way of the eye, proceeding through the visual system to the brain and emerging as comprehended linguistic information. The 'bottom-up' model, due to its hierarchical nature, assumes that any skill deficit at any point in the hierarchy will severely limit reading ability (Ross, 1976).

Top-down models of the reading process are based on the assumption that reading does not require the ability to decode individual words, but rather that reading is a process of sampling text to directly extract meaning. Using cues from the graphic structure of words, the context in which the words occur and the constraints on English language structure, top-down models suggest that the reader can go directly to meaning without the necessity to decode individual words (Goodman, 1971). The theory gathers support from the observation that fluent readers, primarily adults, do not appear to spend sufficient time at any position on a page to completely decode individual words and that adult readers, when reading orally, frequently change words in a text while preserving the meaning of the original word (i.e. 'spoke' for 'said').

The more recent interactive-compensatory model of the reading process is based on the observation that neither the top-down nor bottom-up models satisfactorily describe the reading

process. The interactive compensatory model of reading postulates that, while reading skills may be hierarchical, deficits in one skill can be compensated for by the use of another skill, either higher or lower in the hierarchy. Support for this theory was found in a study that required fluent and poor readers to read sentences which contained words that did not match the context of the sentence, such as 'he fan' for 'he ran' (Allington & Strange, 1978). When asked to read the altered sentence, fluent readers could either read it as written or read it so that it made sense with equal ease. Poor readers, however, were more successful when asked to read the text so that it made sense, that is preserving the context rather than the accuracy of the text. The finding for fluent readers is in accord with the bottom-up models since they could use the graphic cues to determine the actual word and cognitive skills to determine the meaning of the sentence. The bottom-up models cannot explain the finding for the poor readers. Neither finding is in accord with the top-down models since this model assumes that poor readers cannot use context to facilitate word recognition and that fluent readers rarely use graphic cues when reading.

The interactive-compensatory model can successfully explain the finding that poor readers appear to primarily use context cues while fluent readers can use either graphic or context cues with equal skill. This model suggests that, regardless of reading ability, readers can compensate for skills deficits by employing skills from other levels of the reading process. Comprehension requires both grammatical sensitivity and the ability to access knowledge of sentence structure and prior information (Willows & Ryan, 1986). A

child with poor word recognition skills could, therefore, compensate for this deficit by using context and prior information to facilitate the reading process. Although compensation may not be necessary for a fluent reader, the interactive process may still work in situations where full attention to the text is not necessary. The fluent reader can use higher level processes to monitor comprehension and, if attention wanders, the higher level processes will enable the reader to find the portion of the text at which comprehension was first lost. Poor readers do not appear to have the ability to monitor comprehension since the conscious use of context as a facilitory device is used at the expense of other skills in the reading hierarchy (Stanovich, 1980).

The three models emphasize different aspects of the reading process. The 'bottom-up model' emphasizes the visual nature of information extraction, suggesting that visual skills are of primary importance in reading. Applying this model to the reading environment, any external factor which interferes with the ability to utilize visual information, such as degrading the quality of printed material, should ultimately affect comprehension. The 'top-down model' of the reading process emphasizes cognitive skills and would predict that reducing the amount of visual information would only interfere with reading if the degree of degradation was severe enough to remove all graphic cues. The interactive-compensatory model suggests that, although reading skills are hierarchical, this does not preclude skill substitution. Skills such as context prediction could compensate for degraded visual information, however, this may make reading more difficult for readers who rely

primarily on visual skills. Determining how visual skills may affect reading fluency, therefore, is not possible without an examination of the visual skills research.

Visual Skills in Reading

The importance of visual skills in the process of reading was first described by Orton (1928) who found that letter and word reversal, along with mirror writing, were frequently found in children with reading and spelling disabilities. An inability to sustain vigilance has been postulated as a primary reason for such visual disabilities in poor readers (Ross, 1976). This theory was supported in a study using 16 normal and 16 reading disabled boys on a task that required a decision and response to a stimulus letter preceded by a signal letter. The subjects were presented with a series of random letters and were told to respond to the letter X only if it was preceded by the letter A. The results showed that reading level affected both the number of correct responses and the ability to discriminate between true response signals and false alarms. Reading disabled boys responded to significantly fewer true response signals and were significantly less able to discriminate between true response signals and false alarms. There was, however, no difference in the increase in either omission or commission error rates with time for the two groups since all subjects showed a reduction in accuracy as task length increased. This suggests that the reading disabled child has a reduced ability to attend to visual stimuli rather than a simple inability to sustain visual attention (Swanson, 1981).

Theorists who support the psycholinguistic approach to reading, such as Smith (1973), disagree with this finding and contend that fluent readers display a selective attention to visual stimulus and that over-attention to detail is a contributing factor in reading disabilities. A study conducted by Willows (1974) supported the selective attention theory, but not in the manner assumed by Smith. Eight typewritten stories were read by 104 sixth grade boys. The stories read by one half of the subjects included 'distractor' words, which were relevant to the story content, typed in red ink between the lines of the main story. The subjects, who had been classified as either good or poor readers based on the discrepancy between their reading grade level and measured intelligence, were instructed to read the story aloud so that they could answer questions about the content. The experimental group was given additional instructions to ignore the words typed in red since they had been put there as a distraction. When tested later, poor readers made more comprehension errors than good readers, but rarely made errors related to the distractor words. Good readers made fewer errors overall, but the number of errors related to the content of the distractor sentences, or intrusion errors, was higher for the good readers than for the poor readers. Willows speculated that, although good readers do display a selective attention to detail, it is a more global selective attention than exhibited by the poor reader. The poor reader concentrated primarily on the visual aspects of the task, including individual words and lines, therefore stimuli that resulted in intrusion errors for the fluent reader interfered with scanning and decoding for the poor reader. The good

reader, who has automatized the visual skills necessary for reading, concentrated globally on the extraction of meaning, therefore detection of information was restricted to words and phrases that were contextually appropriate.

A study by Schwartz and Stanovich (1981) supported the global selective attention theory of good readers while suggesting that poor readers do not necessarily attend to individual words while reading. Both good and poor readers were asked to read sentences and paragraphs containing misspelled words, such as 'fan' instead of 'ran', that changed the meaning of the sentences. Contrary to the theory of over-attention to individual words, poor readers had more difficulty identifying the anomalous words than did the good readers. In additional testing sessions where the existence of such words was revealed as a treatment condition, the poor readers had more success reading the anomalous words in context than reading them as they were written. The fluent readers, on the other hand, had little difficulty identifying the words in either test condition and could read them either in context or as written with almost equal success.

The results of the studies of selective visual attention demonstrate an interaction between reading ability and task requirements on tasks requiring specific visual skills. Low ability readers find tasks requiring context reading to be easier than tasks requiring sustained attention to the details of words. More fluent readers find either task to be equally easy, however, the ability to perform both tasks equally well reduces the fluent reader's accuracy

when extraneous information is present. The reason for this interaction may be due, in part, to differences in visual scanning ability often observed between good and poor readers.

The ability to scan a line of text and extract information rapidly is a skill used spontaneously by fluent readers (Ross, 1976). In one fixation, the fluent reader can identify four to five words of connected text and can make four fixations a second (Holmes, 1971). In a simple reading task, the poor reader tends to fixate on each word in the text and requires more time to complete a story with less comprehension than does a good reader (Mark, et. al., 1977; Kochnower, et. al., 1983; Willows, 1974). When the amount of information in text is low, that is the words are 'easy', the non-fluent reader can scan at almost the same rate as the fluent reader, however, as the text becomes difficult, the scanning rate of the non-fluent reader drops and word-by-word identification replaces fluent reading (Holmes, 1971).

Efficient text scanning appears to be dependent on the ability to monitor comprehension. Fluent readers appear to monitor their comprehension as the reading process proceeds and to control the input of information by controlling the size and speed of saccades. In cases where the text to be read is unfamiliar, the fluent reader will slow down eye movement and shorten saccade length in order to maximize exposure to unfamiliar words and to extract their meaning from context. Poor readers do not seem to be aware of efficient text scanning behaviour, however, the difficulty seems to be related to their poor reading skills and not to their lack of training in eye

movement. While the fluent reader appears to be able to monitor words in the text ahead of and behind their current eye position, and to actively control visual input, the poor reader does not appear to be aware of the necessity to monitor comprehension while reading (Heiman & Ross, 1974). This theory was supported in one study of visual scanning saccades. When college students were presented with text of varying difficulty level, it was observed that the ability to voluntarily alter saccade length for different task requirements was restricted to fluent readers. Subjects were asked to read passages for either moderate or detailed comprehension or for 'gist'. Only the most fluent readers were able to control their saccade length depending on the requirements of the task. For all subjects, however, the length of saccades interacted with text difficulty, with the shortest saccades being observed with the most difficult text (Ryan, 1981).

The interaction between reading ability and control of saccadic movement also appears to occur in children. Poynter, Schor, Haynes and Hirsch (1982), examined the effect of reading level, as determined by the Stanford Reading Achievement Test, verbal intelligence, as determined by the Peabody Picture Vocabulary Test, and task difficulty on four oculomotor functions in both prose and digit reading conditions. Seventy-four subjects in grades four and six underwent a detailed, near-point optometric examination prior to testing. The forward fixation frequencies, regressive fixation frequencies, average duration of fixations and lag of accommodation were then measured while the subjects read passages at their level of reading ability and pages of random digits.

The data were standardized and regression coefficients were calculated both with and without controls for verbal intelligence.

Verbal intelligence contributed only marginally to lag of accommodation, however, it contributed significantly to the relationship between forward and regressive fixation frequency and reading ability. When verbal intelligence was controlled, none of the individual oculomotor functions contributed significantly to reading ability. When the four functions were combined, however, between twelve and twenty-one percent of the variance in reading ability could be explained by oculomotor functions independent of verbal intelligence. Although the authors admit that they cannot postulate causation, they do suggest that oculomotor functions, primarily forward and regressive fixation frequencies, are significantly related to reading difficulties. When children are reading at their level of ability, however, no one function significantly contributes to reading ability.

Because a reader cannot understand what he cannot decode, and because inability to decode reduces reading speed, comprehension depends on the ability to read accurately, rapidly and fluently. Well established visual attention and visual scanning skills and the ability to monitor comprehension are intimately associated with the ability to read fluently, since performance on tasks requiring these skills interacts with reading levels. The skills, however, appear to be underdeveloped rather than absent entirely, in the non-fluent reader. While the fluent reader is thought to progress through reading stages, beginning with a guessing from context approach,

through a decoding from graphic structure and finally, to a direct extraction of meaning as reading becomes automatic, the poor reader seems to be unable to progress to the automatic reading stage (Stanovich, 1980). However, reading ability interacts with text difficulty. On material that is easy for a child's reading ability level, grammatically and graphically acceptable oral reading errors do not differ for good, average and poor readers. As the difficulty of the reading material increases, good readers demonstrate greater flexibility in reading strategy than do poor and average readers (Blaxall & Willows, 1984). This interaction between reading ability and text difficulty suggests that differences reported in studies of good and poor readers may be due more to the selection of reading material used in the studies than to deficiencies in any specific reading skill.

The research addressing visual and cognitive skills in reading demonstrates that, regardless of the reading model assumed to be operative, no one skill is central to fluent reading. Since children of different reading levels demonstrate differing sensitivities to the requirements of the reading tasks, the ability to use each of the skills, rather than an absence of skills, may determine the degree of reading ability (Blaxall & Willows, 1984; Stanovich, 1980). Consequently, conclusions cannot be drawn from a study that concentrates on only one skill. Examining only one reading skill assumes that the skill selected does not interact with other skills, an assumption that is not supported by the reading literature. Similarly, selecting subjects of only one reading level, or matching subjects on chronological age, assumes that the subjects at all

reading levels will respond to the experimental condition in the same way. Such an assumption is not warranted, as is evident from the reading literature (Backman, Mamen & Ferguson, 1983). Failure to take into account the multiplicity of visual skills and the differing responses of subjects at different levels of reading ability and text difficulty could well lead to erroneous conclusions and may be responsible for some of the confusion seen in the literature related to the use of computers by children.

Since no one visual skill can be identified as being of central importance in the reading process, all visual skills must be considered to be of importance. Any visual factors which may interfere with sustained visual attention, visual scanning, visual accuracy or visual memory must, therefore, either be controlled by the experimental design or measured and accounted for statistically if any conclusions are to be drawn concerning the relative merits of the different media. Since the visual environment obviously can have an effect on the efficiency of visual skills, it is apparent that any study which purports to compare reading from different media must demonstrate that these factors have been taken into account. Effective methodologies to address this important aspect are available in the field of vision research.

Visual Skills and Vision Research

Vision research addresses the issue of how people respond to visual stimuli and how changes in the visual environment can affect such responses. When evaluating reading performance from

different media, such as computer monitors and paper, stimulus differences provided by the different media must be considered. Because a computer monitor includes a glass screen, reflections from light sources such as a window or a room light, can reduce the visibility of screen images. Additionally, until recent technological innovations significantly improved screen quality, the quality of text available on a computer monitor was inferior to that found on paper, a factor which can reduce visual performance (Gould, et.al, 1987). Visual performance can also be reduced by such factors as disability, discomfort or reflected glare and degraded print-background contrast, all of which affect the apparent contrast of the text. An understanding of how such factors can influence the results of a study comparing different reading media is essential for establishing an appropriate experimental design.

Illuminance, Luminance and Brightness

The amount of light falling on a surface is referred to as illuminance and is measured in lumens per meter squared, or lux. Luminance refers to the amount of light reflected from a surface to a detector, such as the eye, measured in candelas per square meter (cd m^{-2}). Since the perceptual counterpart of luminance is brightness (Rea, 1984), luminance measures are most appropriate for studies which are concerned with visual stimuli.

Print-Background Contrast

Luminance measurements are required for determining target-background contrast. Contrast can be defined in terms of luminous

differences between a target and its background with respect to an absolute level of stimulation. Contrast, when defined as differences in the brightness of various portions of the visual field, is the basis of form perception (Carlson, 1981). Since the visual system cannot detect extremely small luminous differences, sufficient contrast is required in the visual scene to provide a stimulus for vision.

One of the earliest English language studies of the effects of reversing contrast on letter identification was carried out in 1931 by Holmes, in an attempt to replicate a study published in Germany by Kirschmann (1908). Letters were printed on cardboard, either black on a white background or white on a black background. Subjects were then asked to identify the letters as they were moved toward them on a sliding frame. In all cases, the black letters were identified faster and more accurately than the white letters. During the same year, Patterson and Tinker (1931) conducted a study designed to test the speed and accuracy of college-level readers using various types of contrast reversal. Readers were asked to read either black on white or white on black words or paragraphs taken from a standardized test with two forms. For all subjects, the white on black arrangement caused a decrease in performance of 10% over all conditions. Methodological flaws reduce the generalizability of this research, however, since the researchers did not specify the lighting conditions used in the study, the distance of the letters from the subject, the clarity of the print or the head positions used in each condition. Changing the background from white to black also changed the absolute level of adaptation. Since a reduced adaptation

level can lower performance, this factor must be taken into account when reversed contrast studies are performed (Rea, 1987).

An ambitious project, using contrast and different types of print was carried out by Taylor (1934). Subjects, again at the college level, were asked to read words, nonsense words, letters or sentences printed in different type sizes in either white on black or black on white. In all conditions, the black on white print proved superior with test condition affecting the absolute scores but not the relative difference between contrast conditions. Type size affected the legibility of both types of print, but black print was identified more quickly than white. The nonsense words were also more rapidly identified for black print than for white print. Familiarity with experimental conditions caused a reduction in the time needed to identify the material, however, as with the other conditions, white print retained a relatively inferior position to black print, regardless of the degree of familiarity. In addition to the simple task of recognizing letters and words, this study also examined reading speed, eye movements and time required for comprehension in the two conditions. As with the recognition tasks, the black print proved to be superior. Smooth eye movements were disrupted with the use of white print and reading speed declined by 11% for sentences and 23% for nonsense words when white print was used.

The lack of specification concerning the test stimuli and test conditions reduces the generalizability of the results obtained in the studies by Holmes (1931), Patterson and Tinker (1931) and Taylor (1934). A similar lack of specification also reduces the usefulness

of a more recent study by Osborne and Holton (1988) in which no reduction in reading speed was reported between computer screens with light or dark characters. In all of these studies, the effect of reversing the usual reading contrast was examined by changing the luminance of the background. In addition to changing the task contrast, this procedure also changes the adaptation luminance which reduces task performance (Rea, 1987). Any changes observed with different task contrasts, therefore, cannot be attributed solely to the contrast reversal unless the effect of different adaptation luminances has been controlled.

The effect of background luminance and contrast on task performance was examined by Rea, Ouellette and Kennedy (1985). Subjects were required to scan juxtaposed columns of numbers for discrepancies. Contrast was manipulated by varying the density of the ink pigment and background luminance was held constant at either 12 or 169 cd m^{-2} . The density of the ink pigment affected both the accuracy with which the task was performed and the position assumed by subjects to perform the task. For columns printed in grey, rather than black ink, subjects assumed postures that were considered uncomfortable in order to achieve a better view of the task materials. Task luminance affected viewing height as subjects moved closer to the task at the lower luminance. Subjective impressions of task difficulty, however, were affected by contrast but not by background luminance.

A recent study by Legge, Rubin and Luebker (1987) found that reading rate was sensitive to contrast reduction primarily for very

large and very small letters. Five subjects were required to read lines of text that 'drifted' across a television screen at an optimal rate of speed for error free reading. Character size and print contrast were systematically varied throughout the testing sessions. Contrasts, which range from 0 to 1.0, were measured using the formula $C = (L_{max} - L_{min}) / (L_{max} + L_{min})$, with L_{max} and L_{min} referring to the minimum and maximum luminances. For a dark print on light background display, L_{max} would refer to the white background and L_{min} would refer to the dark letters. Contrasts for the study were reduced by manipulating the luminance of the dark letters and maintaining the luminance of the background.

For letters that were approximately twice as large as newsprint, reading rate declined slowly from 300 words per minute at maximum contrast to 200 words per minute at a contrast ratio of 0.10. Below this contrast, reading rates declined rapidly, reaching 30 words per minute at a contrast of 0.02. For letters that were extremely large or extremely small, reading rates at maximum contrast were half that of the more standard sized letters and a low tolerance to contrast changes was apparent. A comparison of one subject's responses to contrast changes in static text and drifting text showed that, although reading rate was slower for drifting text, the response to changes in contrast was similar in both conditions. It was concluded that, for readers with good vision, normal reading is tolerant to contrast changes within a wide range of character sizes (Legge, Rudin and Luebker, 1987).

Glare

Contrast can be affected by lighting in one of two ways. When light is reflected from the task to the eye, the physical contrast of a target against a background can be reduced. This is known as reflected glare, or veiling reflections. An example of how reflected glare can reduce contrast can be seen when trying to read text from a glossy magazine in the presence of a bright light reflecting from the page to the eye.

Introducing a bright light source into the field of vision will produce direct, or disability glare. Disability glare reduces apparent contrast due to scattered light within the eye. Although the physical contrast of the target is not reduced, light scatter in the optical media, primarily the crystalline lens, produce a "veil" of light through which a visual target must be detected. Visual performance is most affected by disability glare when the light source is very bright and close to the line of sight of the task (Rea,1984).

Both direct and reflected glare can deleteriously affect visual performance by functionally reducing the contrast of a target. When the contrast of a target is close to the threshold of detection, performance will be affected more than when the task contrast is higher. In many cases, however, the effects of contrast reduction due to both disability and reflected glare can be overcome by manipulating the eye-task-lighting geometry. Subjects in the previously mentioned study by Rea, et.al (1985) actively avoided

reflected glare to improve performance. In some cases, moving the light source, the task or the person can effectively overcome the source of contrast reduction.

Print Quality

The initial processing of text information requires visual attention to a written language code. In order to efficiently process the information, the reader must be capable of rapidly and accurately identifying each word or unit of information presented. A good reader, with normal vision can identify seven digits in a string with only a 10 millisecond exposure time. The contrast of the stimulus letters relative to the background of the presentation medium affects the accuracy of stimulus identification. Several studies, however, have indicated that text can be degraded to a great extent before any decrement is seen in reading performance (Holmes, 1971; Haslerud & Clark, 1957; Holmes, 1931).

Tachistoscopic studies show that a skilled reader can, from a single exposure, identify four to five unrelated letters, two unconnected words or four to five connected words (Holmes, 1971). Since the amount of visual information available in tachistoscopic presentations is much lower than that available in most reading material, it would appear that the reader does not rely solely on visual information for word identification. This assumption has been supported in studies that have examined the effects of reduced visual information on comprehension. Individual letters can be classified according to the number of distinctive features, or

information 'bits', required for identification in isolation (Holmes, 1971). By presenting subjects with words in which successive information 'bits' have been removed, researchers have calculated that only fifty percent of the features of a letter within a word are required for identification of the word. The degradation of the letters, however, required that some distinctive features, such as the upstroke on the letter 'h', were not removed. Providing that such features were distinguishable, readers appeared to be able to recognize words with only fifty percent of the visual information required to identify individual letters (Haslerud & Clark, 1957; Holmes, 1931).

The results of these early studies were supported by Timmers, et. al. (1980), who tested the recognition response time and identification accuracy of three letter Dutch words, presented tachistoscopically to twelve subjects with normal vision. The stimulus words were either typed on white paper through a new carbon ribbon, or degraded by degrees to a contrast level slightly below 50% of the original. Contrast was measured using the formula; $L_p - L_c / L_p$, where L_c is the luminance of the characters and L_p the luminance of the paper. The results of the study showed that when the contrast of the characters was degraded below 50%, recognition of text presented directly to the fovea showed only minor declines, whereas recognition of text presented to the parafovea declined from 10-50%. The authors suggest that this means that the available reading field is reduced when text is not extremely sharp, thus degrading reading recognition skills.

More recently, Smith and Rea (1982), examined the effects of print quality and light levels on the level of comprehension and speed of completion on a standardized reading test. Two groups of subjects, four young adults (average age = 20) and four older males (average age = 54) completed the Davis Reading Test using either the original booklet or a poor quality photocopy. All subjects completed four test sessions under different lighting conditions using alternate forms of the test and alternate print quality. The study showed that poor quality print caused a slight but significant decrement in both speed and level of comprehension but that lighting condition did not appear to affect the results. The authors speculate that the subjects may have been able to overcome the effects of low-level lighting by adjusting their posture to achieve better visibility.

A recent study by Colombo, Kirschbaum and Raitelli (1987) found that, under different levels of luminance, text blur does affect visual performance. Eight subjects were required to search for duplicated words in a list of ten words. The task was repeated for five levels of text blur, defined as the text 'sharpness', at five different luminance levels. The influence of the degree of blur on visual performance was assessed by comparing the time required to identify the repeated words for each word list and each level of adaptation luminance. The authors conclude that the higher degrees of blur, or text sharpness, in combination with very low or very high luminance levels, leads to a decrease in visual performance. The authors suggest that the influence of print quality on visual

efficiency is related to the luminance level of the reading material, however, blur appears to be the more important factor in reducing visual efficiency.

Reading Angle

Printed text is normally viewed at an oblique angle while a computer monitor on a fixed base is viewed at a 90 degree angle. To examine the effect of reading angle on reading speed and accuracy, Gould et. al. (1985), presented eighteen adult volunteers with articles to be proof-read. The papers were presented so that the characters subtended a visual angle of either 6.7, 10.6, 16.0, 24.3, 36.4 or 53.4 degrees. The results of the study showed that, except for the two smallest angles and the largest angle, there was no significant difference in proof-reading accuracy.

The vision literature demonstrates the importance of controlling or accounting for aspects of the visual environment if the results of studies of the reading process are to be generalizable. The visual environment is as important, if not more so, in studies that use computers, rather than paper, as a reading medium. Many computers are equipped with screens that reverse the usual light background-dark print configuration, thus effectively reducing the luminance compared to paper. The presence of a screen covering the reading surface of a computer monitor increases the possibility of task contrast reduction due to reflected glare. Finally, the quality of the image which can be produced by the computer monitor can affect the visual performance of the reader. Any study which addresses

visual performance on tasks presented on a computer monitor must take all of these factors into account.

Reading From Computers and Related Research

The research related to reading and visual skills emphasizes the need to account for different levels of reading ability in any study comparing reading from different media. Since reading level and task performance are not always linearly related (Blaxall & Willows, 1984), studies which address reading performance under different task requirements cannot ignore reading level, nor can they effectively be confined to only one reading level. Although the reading research is primarily confined to studies of children, this requirement could well be extended to include any study of reading performance in which adult subjects are used. Unlike children, whose reading ability can change with maturation, the reading ability of an adult is more static. There are, however, different levels of reading ability within the population of adult readers.

The importance of addressing a variety of reading skills, including comprehension, is also evident from the reading literature. A child with adequate decoding skills may be capable of accurately reading large amounts of text without any understanding. Conversely, a child with good context reading skills may be able to grasp the meaning of a story but may be incapable of reading many of the words. A study which does not include an examination of a variety of reading skills will not fully assess the degree of text comprehension.

Studies from which conclusions are drawn concerning the effectiveness of the computer as a reading medium either explicitly or implicitly use reading performance from paper as a baseline for comparison. The area of vision research indicates that the visual environment can affect the performance of readers. Since a computer monitor differs visually from a sheet of paper, a direct comparison of the two media is required in order to make statements about reading performance from a computer compared to paper.

Studies in the area of vision research indicate that, to properly compare reading performance from a computer monitor to reading performance from paper, the media should be equated for luminance, contrast, character size and image quality. Sources of light which may create direct or reflected glare can result in unintentional contrast reduction. When a computer monitor is used as a reading medium, therefore, the test environment must be illuminated in a manner which excludes light sources directed at the screen or in the readers line of sight.

All of these factors can, to varying degrees, affect the results of studies which address reading. Since the research comparing reading performance from paper to reading performance from computers appears to yield contradictory results, a review of this literature will be undertaken from the perspective of both vision research and reading research to assess the degree to which the various factors have been considered. Due to the different focus of the adult and child literature, the two will be considered separately.

Studies with Adults

The research which has examined the reading performance of adults from computer monitors has been confined mostly to the area of reading speed and accuracy. Only one study, by Hepfner, Anderson, Farstrup, and Weiderman, (1985), has addressed comprehension and none of the studies has adequately considered the effect of different levels of reading ability on reading performance. The research in adults has focused primarily on the technological aspects of computer use and addresses the effects of contrast reversal, image quality and contrast on task performance. Because varied monitor configurations have been used in the literature and standard terminology has not yet been adopted, the terms used most frequently will be maintained.

Standard terms were suggested by The Human Factors Congress in 1980 to describe various monitor configurations. Using photographic terms, a monitor that displays dark characters on a light background is referred to as a 'negative contrast screen' and a monitor with the more usual light characters on a dark background is referred to as a 'positive contrast screen' (Radl, 1980). The more common term for a negative contrast screen, and the one used throughout this paper, is a reversed image screen. The reversed image refers to the page-like display found on the newer, high-resolution screens, produced by reversing the standard light character on dark background display found on most monitors.

Overwhelmingly, most personal computers purchased in the last ten years had a positive image screen with either amber or green characters on a black background. While the reversed image screen has been available for at least five years, the low resolution and excessive flicker of the early displays reduced the amount of interest shown in this type of screen. Although the technology has improved substantially, the cost of the newer, high quality monitors limits their use. Despite the wide distribution of monitors with a dark background, the effect of reading text from a medium that reverses the usual reading format did not receive much attention until 1980. Since then, there has been considerable interest in the area of how reading speed and accuracy are affected by computer monitors of varying configurations. The research has primarily examined the effect of print quality, foreground-background contrast and reading angle on reading speed and accuracy.

To determine whether contrast reversal affected reading skills, Bauer and Cavonius (1980) compared the speed and accuracy of word processing operators using either dark background or reversed image monitors. Based on performance comparisons, proofreading accuracy and subjective ratings, the researchers concluded that the reversed image monitor improved the legibility of the screen and increased the probability that errors would be detected. When identifying nonsense words, the 13 male and 10 female computer operators were reliably more accurate using the reversed image monitor than when a more usual dark background configuration was used. Similar results were found when 12 male

and seven female operators were asked to compare nonsense words typed on paper to words displayed on either a reversed image or positive image computer monitor. Using the reversed image screen improved the detection of discrepancies by eight percent. Although the results are impressive, the study is marred by the fact that each subject adjusted the contrast for personal comfort, thus reducing the degree of standardization necessary for replication.

In a similar study, Radl (1980) found that errors in input and reports of visual discomfort were reduced when operators were asked to type data into a computer equipped with reversed image monitors. Twenty-four computer operators (male and female) were required to enter data into a computer equipped with either a reversed or positive image computer monitor. In addition to a ten percent increase in letters entered per minute, the operators also rated the reversed screen image computer as significantly better in terms of operator comfort. Based on previous studies of screen glare and operator accuracy, Radl suggests that the reversed image monitor, under adequate lighting conditions, reduces the amount of discomfort glare, thus providing more optimal conditions for monitor work. The study was well controlled, however, a definition of 'adequate lighting' would have increased the utility of the study as a methodological model.

Gould, et. al. (1987) examined the effect of lower resolution screens with reversed images on both the reading speed and accuracy of adults who were regular users of computers. Compared to paper, reading speed and accuracy on a proci-reading task was

reduced in all cases. When a similar task was carried out on a high resolution, reversed image monitor, reading speed and accuracy was not reduced compared to paper. Improving the refresh rate of the dark background display slightly improved reading speed and accuracy, however, no one variable could be determined to be responsible for any improvement in reading performance. Any improvement appeared to be the cumulative result of increased refresh rate, improved resolution and reversal of the screen image. The conclusion drawn from this research was that the reduction in reading speed observed in the earlier study was due either to the contrast, the flicker and the quality of the image produced by the computer monitor. Although the computer monitors used in these studies were fully described and a task description was given, there was no attempt made to equate the image on the computer monitor with the image on the paper and therefore, luminance and print contrast for the two media cannot be assessed. Additionally, the study was confounded since both resolution and background luminance were different on the monitors making it difficult to determine which factor was responsible for the observed differences in performance.

Most studies with adults have focused on methods to improve reading speed and accuracy. This focus has meant that the studies primarily compare either the proof-reading rate or the data-input rate of computer operators working at a computer station compared to the rate of the same operators working with paper. Few studies have addressed the process of reading in adults, perhaps because the

process of reading from a computer has been considered to be identical to the process of reading from paper. A study by Zola (1984), which was not directly related to computer use, suggests that computer monitors may, in fact, alter the approach taken to reading, at least in adults. University students scanning computer presented text for anomalous words, showed shortened saccadic jumps, indicating that they fixated on each and every word of the text. The author concluded that the results showed that anomalous words do not interfere with word fixation time since the subjects fixated at each portion of the text equally.

The conclusion reached by Zola (1984) contradicts previous studies of saccadic movements which have demonstrated that fluent readers do not fixate on each word in the text (Smith, 1973, Ross, 1976). What the research did not address, however, was the effect of presenting text on a computer monitor rather than paper. The study did not include a control group of subjects reading from paper, nor did the researcher explain how well the computer generated image compared with a similar paper image. The findings from this study, therefore may mean that previous studies of reading saccades were improperly conducted, as Zola maintained, or may have been due to the quality of the screen image. Alternately, if the screen image was a good approximation of a paper image, the results may have been due to the fact that a computer monitor actually shortens the saccadic jumps normally seen when reading from paper. Because of the methodology employed, it is not possible to interpret the results adequately.

Although Gould et. al. (1987) concluded that any differences observed between computer and paper can be attributed to a reduction in reading speed due to monitor resolution and Zola (1984) reported that saccadic movement differed on computer compared to previous reports based on reading from paper, only one report has found any differences in reading comprehension from a computer monitor compared to paper. Bagnara (1980) determined that, when scanning alone was not sufficient to accomplish a reading task, a reduction in reading speed was accompanied by a reduction in comprehension. Computer operators who were asked to scan paragraphs for spelling errors and repeated words and phrases showed a decline in reading speed on high resolution monitors for repetitions but not for spelling errors. Since proof-reading for spelling errors does not require comprehension, while checking for repeated words and phrases requires an understanding of the intent of the sentence, Bagnara concluded that comprehension of material read from a computer monitor is lower than comprehension from paper.

Only two studies have directly addressed comprehension of material presented on a computer monitor compared to the same material presented on paper. Osborne and Holton (1988), used a procedure similar to that used by Bagnara (1980), however, Hepfner, Anderson, Farstrup, and Weiderman, (1985) used a standardized test to measure comprehension. Although the subjects used were adults rather than children, the results do shed some light on the contradictions apparent in the previous studies. The Nelson-Denny reading test was administered to 85 subjects randomly selected

from a sample of university staff, faculty and students. The subjects were randomly assigned to either the computer first condition or the paper first condition. After twenty minutes, the subjects were asked to stop and exchange with subjects in the other condition. On changing, an alternate form of the test was presented.

The subjects who participated in the study by Hepfner, et. al. were classified as; regular computer users or nonusers, heavy or light television viewers, video game users or non users, heavy or light readers, fast or slow readers and age 22 or under and 23 and over. While significant differences were found between paper and computer scores for some of the groups, the results for raw scores within groups did not reveal any consistent pattern related to reading habits. Although heavy readers obtained significantly higher print than computer scores, fast readers, slow readers and light readers did not differ significantly between media. Similarly, the pattern of scores for computer users and computer nonusers did not show a consistent trend. Both computer users and nonusers, as well as videogame users, obtained significantly higher print scores, while the scores for video games nonusers did not differ between media. Subjects who regularly used computers showed greater differences between media than any of the other subgroups. None of the groups obtained higher scores on the computer version of the test, however, only median scores were reported for each group.

On a retrospective questionnaire, subjects expressed the opinion that reading from print was faster than reading from a computer monitor. The study was, therefore, reanalysed using a

ratio of correct to complete questions to determine whether reading speed may have affected the scores. The analysis showed no difference between scores calculated as a ratio, leading the author to conclude that reading speed was affected, however, comprehension of the material was not affected. Unfortunately, the analysis of the ratio of correct answers to completed answers was performed on the total sample of subjects, rather than on the original subgroups. There was no report of the differences in ratio scores obtained by any of the subgroups, nor the difference in ratio scores obtained by subjects with high versus low print scores. From the data provided, it is not possible to determine whether a change in the ratio of correct to complete answers may have been observed for subjects such as light readers.

There are numerous methodological flaws that reduce the generalizability of this study. The researchers do not indicate the proportion of the test completed prior to the change signal, nor do they indicate the effect of writing both forms of the test in one test session, a procedure that is not recommended by most test developers. Additionally, they give no information on room lighting, contrast between text print and background or monitor print and background. Even accepting that these flaws might not seriously endanger the results, the reduction in number of questions answered still cannot be attributed solely to the computer monitor in this particular study. Since the questions presented on the monitor were actually answered on a computer scoring sheet, the extra visual-motor movements required for this activity may have been at least partly responsible for the reduction in the number of questions

answered. Despite these flaws, there is an indication from this study that subjects of different reading habits may react differently to reading material presented on a computer monitor. Heavy readers obtained higher raw print scores than computer scores while light readers did not differ between media. If the tendency to read is related to reading ability, or if heavy reading improves reading skills, this result could indicate that reading ability interacts with the reading medium to change comprehension scores between media. A more rigorous methodology would be required, however, to determine the effect of reading ability has on reading from a computer.

The research with adults suggests that reading from a computer monitor with low resolution and a dark-background can reduce reading speed and accuracy. The results of all of the studies, with the exception of Osborne and Holton, (1988), indicate that reading performance is superior on a high resolution reversed image screen, however, the effect of computer monitors on reading comprehension cannot be assessed from this literature. Although both Osborne and Holton (1988) and Hepfner, et.al. (1985) suggested that comprehension is not affected significantly, the methodology employed in these studies does not permit any conclusions to be drawn. Bagnara (1980), indicated that comprehension may well be affected by presenting reading material on a computer monitor, however, the method by which he arrived at this conclusion did not include an accepted measure of comprehension.

Although many of the studies with adults include comparisons of reading performance from computers and from paper, most fail to

take into account the necessity to equate the luminance of the paper and the luminance of the computer monitor. In addition, the research rarely gives an adequate indication of the quality of the monitor display compared to the paper. Finally, illumination levels are frequently described as 'adequate' rather than being specified, making it difficult to assess whether the level was 'adequate' for the purposes of the task.

Studies with Children

Information regarding the effects of using computers to present reading material to children is limited. Much of the available information is based on observational studies, conducted without adequate controls, by classroom teachers. The experimental research that has been conducted with children has focused on the effects of computer-assisted learning on language arts skills and details the results of using computer-based learning programs designed to improve the skills of children with academic deficits. Since the focus of this research is to devise methods to improve reading skills in delayed readers, the results do not necessarily pertain to children whose reading skills are not delayed. Because of the limitations of the literature, only a general overview of the current approach to computer use and an examination of those studies which included experimental controls will be presented. None of the research has considered any of the visual factors of importance in comparative studies of reading, consequently, individual studies will not be assessed for adherence to vision research methodology.

Atkinson (1974) conducted one of the earliest reported studies using the computer as a means to teach reading to young children. Using remote terminals, children in five states were able to receive computer assisted instruction in both decoding and reading for comprehension using a computer program designed at Stanford University. An analysis of the effects of the reading program was performed on fifty pairs of kindergarten students matched for age, sex and reading readiness scores, one half of whom received the computer assisted program and one half of whom received classroom instruction. Subjects were given a standardized test both before and after completion of the reading program and relative improvement in scores was compared. The subjects who participated in the computer program gained an average of five additional months in reading scores compared to the control group who received no computer instruction, with boys improving significantly more than girls. After one year, during which no additional instruction was provided, the computer instruction group continued to show a gain of almost five months over those students who did not receive computer instruction.

The researcher concluded from this study that computer assisted instruction can significantly improve the reading scores of elementary children. The conclusion is flawed, however, since there was no control over the effect of removing the experimental group from the classroom and providing a novel reading experience. The control group, who received classroom instruction, did not spend an equivalent amount of time using the computer for non-reading purposes, therefore a comparison between the two groups is of

dubious validity. The within group comparison between sexes is, however, of interest. The male subjects in this study received lower overall scores than the females on all tests. The improvement in reading scores for the male subjects was greater in the computer group than in the non-computer group while females, who received higher scores overall, improved less than half as much as the males. The researcher attributed this finding to the analytic approach to reading required on the computer, an approach which they assumed to favor males. Other studies, such as that by Candy and Edmonds (1982) however, suggest that the improvement may have been due more to the initial reading scores than to superior analytic skills.

The study by Candy and Edmonds (1982), which yielded seemingly inconsistent results, may help explain the discrepancy in scores between the sexes seen in Atkinson's (1974) study. Ten high school subjects, half of whom were experiencing difficulty in spelling, were selected to participate in an experimental language arts program. A computer program was developed to assist the subjects who were having difficulty with language arts, specifically spelling. Both spelling in isolation and spelling in context was included in the program, therefore there was a comprehension component to the tutorial program. The use of the computer as a spelling tutor resulted in a fifty percent reduction in the spelling errors of the low ability subjects after one session. The subjects who required spelling assistance reported that using the computer was an enjoyable experience since they were not chastised when they made errors. Several of the subjects who did not require

assistance reported that the computer appeared arrogant and superior and that entering correct answers was a waste of time since the computer could, at any time, beat them by raising the difficulty level. These subjects did not appear to benefit from the spelling program and some may have been disadvantaged since the subject who received the lowest score during the study was previously assessed as the most able subject in the area of language arts. This subject appeared to find the instructions confusing and was unable to learn how to operate the program.

Although the program developed was an interesting approach to spelling remediation, there was no analysis included in the report beyond the statement that spelling errors were reduced by fifty percent. Consequently, it is difficult to ascertain the usefulness of the program. In addition to the omission of any analysis, there was no description of the computer monitor used, no comparison of the computer program with the results of a standard tutorial approach to spelling remediation and no follow-up studies on the long term effects of the computer-based program.

The procedure used in this study precludes direct conclusions concerning the effects of the computer-assisted learning program on spelling skills. However, the observation that the subjects who received the least benefit from the program were those who did not need assistance supports the finding by Atkinson (1974). The female subjects in Atkinson's study were less in need of assistance than were the males and were also less likely to benefit from the computerized reading program. Both studies, therefore, imply that

subjects who are in need of academic assistance will benefit more from computer assisted instruction than will subjects who do not need additional assistance. The subjects in Atkinson's (1974) study were of kindergarten age while the subjects in Candy and Edmonds (1982) study were in high school, making a direct comparison of the two subject groups impossible, however, the similarities between the more able subjects are apparent. The methodologies employed in the two studies do not permit firm conclusions to be drawn concerning reading ability and reading medium, however.

Papert (1980) developed an approach called 'discovery learning', using Logo as the basic teaching tool, to help children overcome their inherent fear of mathematics. Since computer programming requires a logical, sequential approach, programming in Logo was intended to assist children to think in an organized, logical and sequential manner. In his book 'Mindstorms' Papert claims that children can discover the principles of logical, sequential thinking if they are given the basic rules of programming and then left alone to develop their own learning approach. Numerous anecdotal reports of children developing intricate programs are included in the book, however, there is no firm evidence to support the validity of the discovery learning approach. According to Hawkins (1984) of the Bank Street School who had observed children who had participated in the discovery learning program, these children do no better in their school work than do children who have had no exposure to the program.

Using a different approach to develop logical thinking skills, Riel (1983) developed a writing program for eight children with

delays in language arts skills. The students were encouraged to form teams and write a newsletter that would be suitable for transmission to students in other schools. The computer used a program that provided prompts to assist in organizing and developing the stories and an editing program to correct the final version. After twelve weeks, the length of the stories written on the computer had increased from 53 words to 79 words and words directly related to the activity described in the story had increased from 29 to 63. As the study progressed the students displayed more willingness to edit their final stories, which required the subjects to read the material for comprehension, spelling and grammar errors. Classroom teachers described the students as also being more willing to write in the classroom. There was no change observed in the spelling errors of the students, however, and the grammar and overall quality of the students writing was not evaluated. Although the students reported that they enjoyed computer-based writing, an important facet of any educational experience, the academic utility of the program cannot be determined from the study as reported.

A recent Australian study assessed the effectiveness of an experimental program, which provided verbal feedback on accuracy in a reading skills exercise. To improve the reading skills of reading delayed children, Grocke (1983) tutored a group of seven children who were required to read paragraphs presented on a computer and choose missing words from a multiple choice selection. A control group of seven children were exposed to the standard classroom curriculum, which included withdrawal for reading tutorials. The

subjects, who were matched for age and reading ability, spent the same amount of time on the computer, but only the experimental group were exposed to the computer based reading program. Pre and post tests were conducted using an oral reading test of basic words. The subjects who were exposed to the computer based reading program showed significant, sustained gains in oral reading after five hours of tutoring, compared to subjects who received the standard curriculum. The results of the study are impressive and the documentation provided concerning the computer program and required equipment is extensive. There is, however, no indication as to whether the gains in oral reading scores were transferred to the classroom. Nor is there a direct comparison between the program material presented on computer and a similar program presented using standard text.

The studies conducted by Candy and Edmonds (1980), Riel (1983) and Grocke (1983) concentrated on the effect of computer based instruction on the performance of academically delayed children. These studies did not compare the performance of academically delayed subjects to a control group of average subjects, nor did they compare the subjects' performance on a similar task using pencil and paper. Additionally, the subjects exposed to the computer in these studies were removed from their classrooms into a novel environment and a novel task medium was used. The addition of a novelty variable to the studies means that it may not be possible to attribute any change in performance solely to the computer. The effect of novelty on performance was apparent in

one comparison of the spelling performance of children given computer-based practice and children given rewards for spelling improvement without practice (Torgeson and Young, 1983). In some subjects, spelling scores improved significantly without computer practice when incentives were offered. Therefore, changes in performance in many of the studies may have been due, in part, to factors such as novelty.

The Computer and Reading Comprehension

Although the studies with children compound the confusion which exists concerning performance on computers, there are some indications that presenting material on a computer monitor is not the same as presenting similar material on paper. The research by Grocke (1985) and Candy and Edmonds (1980) implies that the computer can be of assistance to children with academic difficulties, however, the designs used in these studies did not include adequate controls. Both Atkinson (1974) and Candy and Edmonds (1980) also reported that the improvement in scores on the computer occurred most frequently in those subjects who required assistance, suggesting that performance on the computer may be dependant on the subjects' academic level. Although both studies neglected to include a paper and computer task comparison, they do indicate that reading from a computer monitor may affect comprehension, especially when considered in conjunction with several studies which appear in the adult literature.

Hawkins (1983) and Tucker (1983) examined the overall performance of children regardless of academic abilities and concluded that there was no evidence to assume that the computer had any effect on academic performance. Similarly, based on an overall comparison of the percent of completed questions answered correctly for all subjects on computer and on paper, Hepfner, et.al. (1985) concluded that the computer did not affect comprehension in adults. However, Atkinson (1974) and Candy and Edmonds (1980) examined the performance of subjects at different levels of academic ability and concluded that the computer differentially affected performance. When the raw scores obtained by subjects in the study conducted by Hepfner, et. al. (1985) were analyzed by group, a similar conclusion was reached, with the degree of interference exerted by the computer interacting with reading habits and the degree of computer familiarity.

Examined from this perspective, it is possible that the apparent contradictions in the literature may be due to a combination of indiscriminate subject selection and inappropriate methodologies. The reading literature indicates that the level of reading ability interacts with performance on some reading tasks. This suggests that the seemingly inconsistent results obtained from some computer-based studies may have been due to an interaction between reading level and task. However, since these studies either failed to include a direct comparison between paper and computer, or ignored the importance of reading level, any interactions which

may have been responsible for the results would not have been observed.

Based on the established methodology of vision research and the approach taken to studies of the effect of computer presented text on adult reading skills, any study designed to measure the effect of presenting material on a computer compared to paper must take into account the differences between the media. Unlike paper, a reversed image computer monitor is illuminated by a background phosphor. Although this difference cannot be controlled, attention to the luminances of the two presentation media can reduce the effects of the different lighting configurations. By measuring the luminances of the two media and by taking into account the subjective assessments of representative subjects, a close approximation of equality between media can be achieved. Given equality, a direct comparison between paper and computer monitor can then be made.

In addition to the luminance of the media, room illumination must be considered. Few studies, especially those using children as subjects, adequately describe the lighting conditions under which the studies were conducted. Direct light shining on the computer screen, such as from windows, should be avoided since the possibility of reducing visibility through reflected glare is a factor which must be considered. Without accounting for this factor, any results which report a decrease in reading performance from computer monitors are suspect.

Any examination of the effect of computer monitors on the reading ability of children must take into account the difference between an adult reader with years of experience and a child still in the process of skill development. While adult studies may, arguably, be restricted to measures of interference in established skills such as reading speed, a study that uses children as subjects must be specifically designed to include a measure of comprehension in addition to a measure of reading speed. If a valid measure of comprehension is to be obtained, a test of known validity is necessary. Since children learn to decode text before learning how to go directly to comprehension (Pidgeon, 1976), the test must include a measure of both literal and inferential comprehension. A measure of literal comprehension is necessary to determine whether a child has the skills to search for and retrieve information from a passage of text. A measure of inferential comprehension is necessary to determine whether a child can synthesize the information presented with information acquired in the past in order to answer questions which do not have readily apparent answers. Given the variability of children during the development of the reading process (Schwartz and Stanovich, 1981), a test of reading speed is required to gain an accurate measure of reading rate as it applies to reading fluency. A measure of reading speed or reading comprehension in isolation would not be sufficient to determine the effect of presenting reading material on computer monitors to children with undeveloped reading skills. The Stanford Diagnostic Reading Test (SDRT), with separately standardized subtests of comprehension and reading speed is one of the few tests which

meets these criteria. Although the test has been criticized for its reliance on reading speed as a measure of reading grade level (Tierney, 1976), the validity and reliability of the individual subtests is accepted (Salvia & Ysseldyke, 1985). Test-retest reliability for the various comprehension subtests are reported to be between 0.82 and 0.94, while test-retest reliability for the reading speed subtest is reported to be 0.78.

Using a test with known validity to examine the level of comprehension from computer monitors is not sufficient if the study is not designed to directly compare the level of comprehension of material read from paper and material read from computer. Any study that does not include a direct paper-computer comparison cannot presume to draw any firm conclusions concerning reading comprehension since comparisons of standardized test scores to other measures of reading ability, such as teacher assessments or classroom scores, are not necessarily accurate. Many studies cited as proof that the computer can be used as a reading tutor or as a spelling aid do not include such comparisons, therefore the results are open to the criticism that experimenter attention or other extraneous variables, and not the computer, may be responsible for apparent changes in ability scores. Only a direct comparison of the scores achieved on a paper version of a standardized test with those achieved on a similarly formatted, alternate form computer version of the same test can fully address these issues.

CHAPTER 2

Description of the Present Study

Assessment of Comprehension: To determine whether comprehension differed on computer monitors compared to paper, three studies were designed which would measure comprehension in each of three reading levels. The test instrument selected for this study was the Stanford Diagnostic Reading Test (SDRT) which includes questions which address both literal and inferential comprehension and which provides a separate subtest of silent reading speed and fluency, assessed by means of context prediction ability. Four levels of the SDRT are available which enables children of different grade levels to write a complete test at their own reading level rather than selected portions of a larger test. Additionally, each subtest of the SDRT is separately standardized, allowing the subtests to be administered and scored in isolation. The selection of subjects for participation in this study was restricted to grades four and five. Children in these two grades were young enough to still require grouping for reading instruction, making teacher assessment of reading level feasible, yet most would be advanced enough to have acquired the basic reading skills. Subjects were assigned to one of three reading level groups based on the teachers' assessment of reading skills. A repeated measures analysis of paper and computer comprehension scores could then be performed separately on each of the three subject groups.

The use of two grade levels meant that two levels of the test were required for this study, one intended for grade four readers and

one intended for grade five readers. Because the grade four and grade five tests were of different lengths, raw scores for subjects in the two grades could not be analyzed together. To permit analysis of the combined scores, raw scores were standardized separately for each grade and analyses were performed on standardized scores [z-scores] (Appendix 1).

Calculation of Comprehension Scores: The scoring method suggested for the SDRT uses the number of questions completed correctly to determine reading grade level. Because the test is constructed so that few, if any, children can complete the entire test, individual reading speeds heavily influence reading grade level and a slow, careful reader who correctly completes a small portion of the test can be penalized (Wallace & Larsen, 1979). Although this would not be a factor in a repeated measures comparison, it would influence the between subjects comparisons for the effects of reading level. Two methods were available to overcome this reliance on individual reading speeds; allowing the subjects to complete the test regardless of the amount of time needed, or calculating the proportion of correct answers compared to completed answers for each subject as proposed by Hepfner et. al. (1985), and then examining the effect of medium on the number of questions completed. To avoid the possibility that some subjects may have required extremely long time periods to complete the test, the second solution was selected.

Design Considerations: The study was designed to account for as many extraneous visual variables as possible. Three different

computer monitors were used in the study. To compare the effect of resolution on comprehension scores, each of the monitors used had a different resolution. One, a Zenith monitor running on an Apple IIe, had a dark background with green phosphor text while the other two monitors were high resolution Apple monitors with a light background and dark text. The three monitors also refreshed at different rates (Appendix 2). The luminances of the computer generated images on the high resolution monitors and the paper test sheets was determined prior to testing. As far as possible, the computer monitor and the paper test sheets were equated and no changes were made to the computer monitor throughout the study. For the high resolution monitors, characters on a printed test sheet, when laid over the computer generated image, appeared to be identical in size and print density. It was not possible to control for the fact that the high resolution monitors were illuminated by a background phosphor while the paper was illuminated from overhead light, however, the Zenith monitor had a dark background, allowing some inferences to be made from a comparison of the results from the three monitors.

Illumination levels in the rooms assigned for testing purposes were measured prior to beginning the study and, fortunately, the measurements were similar. One test room included a window at the extreme end of the room; therefore, the test area was arranged to be away from the window and the window was shaded during the test sessions. All test sessions were conducted in the same rooms

and overhead fluorescent fixtures were the only source of ambient illumination.

Due to the scarcity of reports which have included children of differing levels of reading ability, a preliminary study was necessary to test the hypothesis that differences in the comprehension scores of children of different reading levels could be detected between media and to estimate the magnitude of change which could be expected. Additionally, a preliminary study was necessary to calculate the number of subjects needed to demonstrate the difference. A repeated measures design was used in which subjects assessed by their classroom teachers as belonging to one of three reading levels completed a modified version of the comprehension subtest of the SDRT and an oral reading speed test, on both computer and paper. The study was designed so that the order of presentation of the test on the two media was balanced within reading levels and the alternate form presentation order of the test was balanced across media and within reading levels (Fig 1). Only one computer monitor, similar to that found in most classrooms, was used in the preliminary study, so a comparison of resolutions was not possible in this study.

Study 2 included a direct comparison of scores obtained on alternate forms of the SDRT, presented on both paper and computer, using the balanced design shown in Fig. 1 to account for test presentation order and differences between forms. Two computer monitors of differing resolutions were used in this study permitting an additional analysis of the effect of resolution on comprehension scores. Study 2 also included a control group of subjects who used a

keyboard to input answers on both the paper and computer version of the test. The procedure used to test this group was identical to the procedure shown in Fig 1. A variable which could not be controlled was the angle of the computer monitor compared to paper. Some evidence exists that angle does not influence reading speed or accuracy in adults (Gould, 1987). Since computer monitors are normally viewed at an angle different to that of paper, a decision was taken to test the reading speed and accuracy of several children reading from paper presented both horizontally and vertically. No difference was observed, therefore the report on adult behaviour was accepted for the purposes of this study. The results are reported in Appendix 3.

The difficulty of reading material is known to interact with reading ability (Blaxall & Willows, 1984), however, in Study 1 and Study 2, only comprehension scores and not reading errors were of interest. Since the difficulty of the reading material might have interacted with both reading ability and the reading medium to affect comprehension, two levels of test difficulty were used in Study 3 to examine this possibility. Test difficulty was manipulated by administering the grade four test to average and below average grade five subjects. Because the grade five comprehension subtest included a wider range of grade equivalents than the grade four test, the level of difficulty of the questions was considered too advanced for the grade four subjects. Rather than manipulate the difficulty level by also administering the grade five comprehension subtest to the grade four subjects, an alternative approach was used. The grade four test did not include a silent reading speed subtest,

Figure 1: Design Schematic for Studies 1 and 2. Reading level assignment was based on teacher assessment of reading ability. Within reading levels, medium presentation order, test form and test form order were balanced.

therefore all grade four subjects were given the grade five silent reading speed subtest during the test sessions. Since the context prediction portion of this subtest required comprehension of the material, an analysis of the effect of increased difficulty on comprehension scores could be performed.

Analysis of Scores: Subject assignment to one of the three reading groups used throughout the study was determined by teacher assessment prior to testing. Each reading level, therefore, constituted a discrete group and could be analyzed separately. All analyses were performed using the multiple groups analysis program included in the statistics package Systat® version 3.1. Because of the multiple comparisons performed within each of the groups in Study 2, the significance level ($p \leq 0.05$) was divided by the number of comparisons performed to reduce the probability that some of the comparisons might be significant by chance.

Standardized (z) scores were used for all analyses performed on combined grade four and five scores. Proportion scores (correct answers/complete answers) were used for analyses performed within a single grade. Means and standard deviations for all groups on all variables are reported in Appendix 1.

STUDY 1

Method

Subjects:

Following approval of the study by the Carleton School Board, the principals of three schools were approached and permission was granted to send consent forms home to the parents of children in the grade four and five classrooms in the three schools (Appendix 4). Subjects for both Study 1 and Study 2 were recruited at this time. Two hundred consent forms were distributed, of which one hundred and fifty were returned with ten refusals.

Following return of the consent forms, the grade four and five teachers in the three schools were approached and asked to assist in assessing the reading level of the participating children in their classrooms. The teachers were asked to use the following criteria:

Above average reading skills: A student in grade four or five who is reading approximately six months above his or her grade level; i.e., in April, a grade five student who is reading at a mid-grade six level.

Average reading skills: A student in grade four or five who is reading within five months of his or her reading level; i.e., in April, a grade five student who is reading a late grade five level.

Below average reading skills: A student in grade four or five who is reading six months or more below his or her grade level; i.e., in

April, a grade five student who is reading at a late grade four or early grade five level.

In each grade, the subjects were divided into the three reading levels based on the teacher's assessments and subjects for each study were randomly drawn from the three groups. For Study 1, twenty-nine children were drawn from the selection of subjects whose parents had signified their consent. The average and below average groups consisted of eight grade five and four grade four subjects. An above average group of twelve subjects was initially intended, however, after testing five above average subjects, all but one had completed the entire test in less than the allotted time, indicating that a 'ceiling effect' occurred. The five subjects tested were retained for comparison purposes and the remainder of the above average subjects were retained for later study.

Apparatus:

The preliminary study was designed to examine comprehension in children reading from computer monitors that closely approximated those currently used in local schools. The computer used to conduct the study was an Apple IIe with a fixed keyboard and a green-on-black monochrome monitor. The monitor was capable of displaying eighty columns and twenty-five lines per screen. It was possible to display the reading paragraphs and at least one of the test questions on the screen using the word processing program Magic Windows®. It was not possible, however to automate the question answering format, nor to include a measure of silent

reading speed, due to the limited programming capability of the computer. Although not an ideal compromise, a measure of oral reading speed and accuracy was included.

Test Instrument:

Two modified forms of the Stanford Diagnostic Reading Test were constructed by selecting subsets of questions from the reading comprehension subtests of the two forms of the SDRT. Each form of the reading comprehension subtest contained eight paragraphs and either 48 questions, for the grade four test, or 60 questions, for the grade five test. One half of the test questions could be answered directly from information contained in the reading paragraph (i.e., What is the name of Joe's dog?) and were designed to measure literal comprehension abilities. The other half of the questions required the subjects to go beyond the information in the paragraph (What did Mary mean by...?) and were designed to measure inferential comprehension.

To construct the modified forms of the test required for Study 1, the eight paragraphs were divided, consecutively, into four pairs and for each form, one paragraph was selected from each pair so that the length of the paragraphs and the number of questions addressing each paragraph were comparable across forms. The final forms of the modified test, labelled form A and form B for identification purposes, each contained a total of twenty-five comprehension questions for the grade five version and twenty questions for the grade four version.

Using the word processing program Magic Windows[®], the paragraphs and corresponding questions were entered into an Apple IIe computer equipped with a Zenith monochrome monitor. Printed versions were prepared on a dot matrix printer using the same type face as that displayed by the computer monitor. The computer version of the test comprised four complete screens of text which included the reading paragraphs and at least one question per screen. Questions which would not fit on the visible area of the screen were continued on the screen immediately following. On completion of the questions which appeared on the screen containing the reading paragraphs, the area of the screen directly below the paragraph was cleared by the experimenter which brought the remaining questions into view so that the reading paragraph always appeared immediately above the questions. The paper version of the reading test comprised four sheets of individual reading paragraphs with questions and answers arranged directly below the paragraphs. Answer selections were circled by the subject.

The four keys which were at the approximate center of the keyboard (f,g,h,j), were coded as A,B,C,and D to correspond to the answer selections for each of the questions. Key presses by the subjects were recorded on paper by the experimenter since the technical limitations of the equipment precluded computer storage of answer selections.

Task Description

The subjects were required to read short, descriptive passages from either the paper version or the computer version of the test and answer comprehension questions related to the reading passage. The paper version of the test required the subjects to circle the appropriate answer, which appeared as one of four choices directly below the question. The computer version of the test required the subject to press one of the four labelled keys on the computer keyboard which corresponded to the selection of answers.

Procedure

Comprehension: The subjects were tested in pairs in a room assigned by the individual schools. Each pair of children was taken from their classroom to the testing room and placed at opposite ends of a library table. Subjects who were assigned to the paper test condition were given four sheets of paper, stapled at the top left corner, which contained the reading paragraphs and test questions. Subjects assigned to the computer condition were placed in front of the computer and allowed to assume a comfortable position. Stacking chairs were used for seating so that subjects who were shorter than average could be seated on two stacked chairs. Viewing distance from the monitor was not fixed.

The directions for completing the task were given to both subjects, verbally, by the experimenter, with the directions for the answer format being given to the subjects separately. The subject

assigned to the paper version was questioned to determine whether any confusion existed as to the directions. When the subject indicated that the directions were completely understood, the experimenter took a seat close to the computer and the test session was begun.

Because of programming limitations of the Apple 11e, the experimenter sat with the subject assigned to the computer and, when the answer selection for each test question was completed, scrolled the next question to a position directly below the reading paragraph. The entire process was repeated for each subject on the following day when the alternate version of the test was written.

Oral Reading Speed and Accuracy: Following completion of the comprehension portion of the test, the grade five subjects were asked to read a paragraph orally from another portion of the test, either from the computer monitor or the paper test booklet depending on test condition. The reading session was taped for later analysis. A test of oral reading was not conducted in grade four subjects due to the fact that the grade four reading paragraphs were too short to provide a meaningful measure of oral reading speed and accuracy.

Results

Since the subtest version used in the study did not represent the complete comprehension subtest of the SDRT with a balanced number of literal and inferential questions, only total comprehension scores were calculated. To obtain a measure of written comprehension, a ratio of correct answers to the total number of test questions completed was calculated for each subject. To account for differences in total number of questions for each grade level, the scores for subjects in grade five and subjects in grade four were standardized separately and then combined for the final analysis. The oral reading speed score represented the average number of words per second and was calculated by the formula; $(\text{number of words}/\text{time required}) \times 60$. Oral reading accuracy was determined by calculating the ratio of the number of words read correctly to the total number of words read.

To test the null hypothesis that there would be no difference in the scores obtained on a paper version of the SDRT comprehension subtest and the scores obtained on a computer version of the SDRT comprehension subtest for any of the three reading levels, it was first necessary to determine if the three reading levels differed on the paper version of the test. A between groups multivariate analysis of variance was performed on the scores from the paper version of the test for each of the variables to determine whether subjects in the three reading levels differed significantly on comprehension, oral reading rate and oral reading accuracy scores. The results of the analysis are shown in Table 1.

Table 1

Test for Differences Between Reading Levels On Oral Reading Rate, Oral Reading Accuracy and Total Comprehension Standardized Scores on Stanford Diagnostic Reading Test Questions.

Variable	Source	Sum of Squares	DF	Mean Square	F-Ratio
Reading Rate ¹	Level	1.288	2	0.644	< 1.0
	Error	17.712	17	1.042	
Reading Accuracy ¹	Level	4.979	2	2.490	3.019
	Error	14.021	17	0.825	
Total ²	Level	20.042	2	10.021	42.050*
	Error	5.958	25**	0.238	

* $p \leq 0.01$

¹ Grade five subjects only

² Grade four and five subjects

* one grade five subject was not available for the second test

There was a significant effect for level on total comprehension scores on the paper version of the test. Post-hoc comparisons showed that the difference between reading levels was due to a difference between the average and below average reading levels ($p \leq 0.000$). There was no difference between the average and above average reading levels ($p \leq 0.075$), however, as noted earlier, there were only five subjects in this group. There was no difference in the oral reading rate scores or the oral reading accuracy scores between the three reading levels. The modified paper version of the Stanford Diagnostic Reading Test was able to discriminate two of the three reading levels for total comprehension. A univariate analysis of variance was then performed to determine whether

TABLE 2.

Within Reading Levels Contrast Of Paper-Computer Standardized Scores on Stanford Diagnostic Reading Test Questions in Grade 4 and 5 Subjects

Variable	Reading Level	Contrast	Mean	DF	F-Ratio
TOTAL	Above Average	Paper	1.053	1,4	< 1.0
		Computer	1.113		
	Average	Paper	0.565	1,10	26.710*
		Computer	-0.627		
	Below Average	Paper	-0.957	1,11	34.766*
		Computer	0.111		

* $p \leq 0.01$

paper and computer scores differed within each of the three reading levels. The results are shown in Table 2.

Subjects assessed as above average showed no difference in scores on computer compared to paper ($p \leq 0.925$). Both average and below average subjects differed on computer compared to paper, however, the below average subjects obtained higher comprehension scores on the computer version of the test (47% compared to 65.9%; $p \leq 0.001$) while the average subjects obtained lower scores on the computer version of the test (74% compared to 60.8% $p \leq 0.025$) (Fig 2).

Figure2: Mean Standardized Scores Obtained on the Paper and Computer Modified Versions of the Comprehension Subtest of the Stanford Diagnostic Reading Test. Mean Scores represent the ratio of questions correct to questions completed. An asterisk identifies mean scores on the computer version of the SDRT which differ significantly from scores obtained on the paper version of the SDRT (n= 5 [above average]; 12 [average]; 12 [below average]; $p \leq 0.001$).

An analysis by reading level of the number of questions completed on computer compared to paper by subjects in both grades showed that the increase in scores could not be attributed to a reduction in question completion (paper mean=20.75; computer mean=19.9: $F=0.972$, $p \leq 0.337$).

The teacher's assessment of the subject's reading level appeared, from the results of the analysis, to be accurate. Since further studies would also rely on teacher assessments, a discriminant analysis of scores by reading level was performed and the centroids obtained were used to predict group membership as a verification of the accuracy of the teacher's assessment. The results of this analysis are shown in Table 3.

TABLE 3.

**Comparison of Teacher's Assignment and Predicted Assignment
to
Reading Level**

		Predicted Reading Level			Total
		1	2	3	
Assigned	1	5	0	0	5
Reading	2	2	9	0	11
Level	3	1	0	11	12
TOTAL		8	9	11	28

The teachers assessment of reading level matched the predicted reading level twenty five times out of twenty eight for an accuracy rate of 90%. The accuracy rate, although high, suggested that further studies should include extra subjects and that discriminant analysis and tests for outliers would need to be performed.

Summary

The results of the preliminary study led to the rejection of the null hypothesis that the scores obtained on the paper and computer versions of the modified SDRT subtests would not differ. A significant difference in comprehension scores was observed in children assessed as average or below average readers when the results of comprehension questions answered on paper were compared to the results of similar comprehension questions answered on computer. The degree of change was large enough to be observed in twelve subjects and the direction of the change for each of the subjects in the two groups was the same for all but one subject. The increase in comprehension scores for the below average readers and the similar scores between media for the above average readers, however, was unexpected. The results for these groups strongly suggested that the computer did not interfere with good readers but was a beneficial medium for below average readers. For average readers, the computer appeared to be disadvantageous.

The procedure used for the preliminary study was not sufficiently rigorous to draw any firm conclusions, however. The computer used in this study had a dark background with light text, a configuration that has been shown to reduce reading accuracy on computer in adults (Gould, 1985). The length and difficulty of the test did not provide an accurate assessment of the skill level of the above average readers and the test questions represented only a portion of the complete test resulting in a more rapid increase in the difficulty of the test material than is found in the complete SDRT subtest. Additionally, the answer format for the computer version of the test required the experimenter to be in attendance at all times. Since below average subjects frequently perform better in a one-to-one teaching situation (Dunn, 1973), this may have had a positive effect the scores of the below average readers. Finally, an oral reading speed test was used rather than the silent reading speed subtest from the SDRT since the computer used was not suitable for the silent reading speed subtest. The results did indicate, however, that there is a difference in comprehension when reading from a computer monitor compared to reading similar text from paper and that reading level does interact with the reading medium. Since the difference in performance may have been due to differences in the visual characteristics of the two media, a study in which the media were closely equated with respect to size, contrast and luminance was undertaken.

CHAPTER 3

STUDY 2

Study 2 maintained the repeated measures balanced design of the preliminary study, however, the extraneous variables introduced by the equipment used in Study 1 were better controlled in this study. The computer monitors used produced an image which closely resembled the paper image in size, contrast and print quality. Luminance values of the two monitors and the paper test sheets were measured prior to testing. Contact between the experimenter and subjects was minimized by means of an automated answering program.

The modified versions of the test used in Study 2 included the silent reading speed subtest and the complete comprehension subtest of the SDRT. The computer monitors used in this study could display a full test page, therefore, the reading paragraph, test questions and multiple choice answers could be displayed on one screen. A pencil was used to answer questions on the paper version of the test, however, an additional group of subjects, who answered the paper version of the test using a keyboard, was also tested to determine whether the response medium produced any changes in comprehension scores.

PART 1: Reading Comprehension

Method

Subjects:

Forty-eight grade five and twenty-four grade four children were drawn from the remaining pool of subjects whose parents had signified their consent to participation in the study and whose reading level had been assessed by the classroom teacher. As far as possible, the groups were constituted to contain even proportions of male and female subjects.

Teachers were asked to exclude from the consideration children who: had vision that could not be corrected to normal; required special services due to global learning handicaps; suffered from a physical handicap that would interfere with use of a keyboard; or demonstrated a degree of spelling impairment sufficient to require extensive remediation. Since these factors required knowledge of special services requirements, it was not possible to determine whether any children were excluded for any of these reasons. Children who had no experience or extensive experience with computers, as determined by a post-test interview, were also to be excluded from the study. Since none of the subjects reported extensive experience with computers and all subjects were exposed to computers in the school, however, no exclusions were required.

Apparatus:

The reading comprehension subtests from form G and form H of the Stanford Diagnostic Reading Test were entered into Excel[®], a

spreadsheet program, on a Macintosh II computer. Full screen individual spreadsheets, which were displayed without gridlines, were divided into two wide columns and one narrow column. The reading paragraphs were typed in the wide left hand column and the answers were typed into the wide right hand column. The narrow column, which appeared on the extreme right, contained spreadsheet cells which were displayed with outlines and into which the answer selection was entered. A sample test page, containing a question similar to that used in the study, but not taken from the SDRT, is shown in Fig 3.

Text entered into either of the wide columns was protected, using the 'protect document' capability of the Excel® program, to prevent accidental typeovers during test sessions and to facilitate answer entry. The answer cells in the narrow right hand column were individually unprotected and only those cells would accept typing. Each spreadsheet was saved with the cursor positioned on the first answer cell so that the test pages appeared with the cursor in the correct position to enter the first answer.

The sets of spreadsheets containing the test questions were duplicated and stored in individual folders for each subject. For each test session, the complete set of spreadsheets, including a cover sheet to prevent the subjects from being exposed to the test before instructions were completed, was loaded to the Excel® desktop in

Figure 3: Sample Test Page Layout for the Computer Version of the Stanford Diagnostic Reading Test. NOTE - The test page reproduced here does not contain material found in the Stanford Diagnostic Reading Test and is intended to illustrate only the presentation format of the test.

File Edit Format Data Options Macro Window	
Example	
<p>Beautiful, wonderful sights to see And wonderful sounds to hear: The world is the place for a seeing eye And a place for a listening ear.</p> <p>Puppies and lambs and kittens to touch! Softies and silks to feel! Sugar and salt and honey to taste! Fragrant fruits to peel!</p> <p>Beautiful, wonderful, pleasant world! And a child who would know it well Has everything to see and hear, To touch and taste and smell.</p> <p style="text-align: right;">James S. Tippel</p>	<p>A. The poem is intended to--</p> <ol style="list-style-type: none"> 1) Give information 2) Give a description 3) Explain how the world appears to the blind 4) Tell how animals feel <p>B. Kittens and lambs are -</p> <ol style="list-style-type: none"> 1) beautiful 2) helpful 3) to touch 4) to hear
<p>C. The poet thinks that the world is--</p> <ol style="list-style-type: none"> 1) sad 2) cold 3) wonderful 4) scary <p>D. The poet thinks the world is a place far--</p> <ol style="list-style-type: none"> 1) The mind 2) Animals 3) The senses 4) Eating <p>E. The poet describes fruit as--</p> <ol style="list-style-type: none"> 1) Tasty 2) Sour 3) Sweet smelling 4) Unpleasant 	<p>A</p> <p>B</p> <p>C</p> <p>D</p> <p>E</p>

sequence. The Return key of each computer was programmed, using QuicKeys ® to activate an Excel® macro. Using the macro, the Return key, which was labelled 'Next Page' automatically saved and closed the top spreadsheet, leaving the next test page available for reading.

Four keys on the numeric key pad portion of the detachable keyboard, were labelled as 1,2,3 and 4, using red stickers to enhance visibility. The Return key, as previously mentioned, was labelled with a red sticker to denote 'Next Page'. The Enter key on the numeric key pad was labelled as a 'Go' key. Subjects entered their answers by selecting the appropriate key from the numeric key pad and then pressing the 'Go' key. The cell and document protection used to prepare each test sheet, described previously, caused the cursor to advance, incrementally, to the next unprotected cell in the spreadsheet when the 'Go' key was pressed (Appendix 5).

Two computers, one of which was equipped with a Sony RGB color monitor and the other with a Sony high resolution monochrome monitor, were configured as described. The monochrome monitor, with a 'good resolution' was the standard Sony RGB monochrome monitor provided with the Apple MacIntosh II computer, with a vertical scan rate of 67 Hz and a screen resolution of 76 dots per inch. The color monitor, with a 'high resolution' was a Sony RGB color monitor with an approximate vertical scan rate of 65 Hz and a screen resolution of 72 dots per inch. Both monitors were set to achieve maximum clarity by using a combination of subjective judgements by elementary school-age children, who rated the

computer screens for ease of reading and similarity to text, and by luminance measurements of both the computer versions and printed versions of the test. The final monitor settings were maintained throughout the study.

Both the computer and paper versions of the completed comprehension subtest contained eight reading paragraphs with a total of sixty questions for the grade five subjects and forty-eight questions for the grade four subjects. For each grade, one-half of the questions represented literal comprehension questions and one-half represented inferential comprehension questions. The paper version of the test was prepared by printing the spreadsheets sideways on a laser printer, using the same font and test format as that displayed on the computer monitor (Fig 4).

The silent reading speed subtest was constructed in the same manner as the comprehension subtest. The subtest contained two elements, a reading speed measure, which was defined as the absolute number of words read in three minutes, determined by the last answer entered in the answer boxes and a fluency measure, which required the subjects to predict, from context the correct answers to sentence completion questions. Since this is not the usual definition of fluency, the subtest is referred to as 'context prediction' throughout the study.

The answers entered by the subjects on the computer version of the test were scored automatically, using a marking program written in Excel®. Answers entered on the paper version of the test

Figure 4: Sample Test Page Layout for the Paper Version of the Stanford Diagnostic Reading Test (reduced to 90% of original). NOTE - The test page reproduced here does not contain material found in the Stanford Diagnostic Reading Test and is intended to illustrate only the presentation format of the test.

<p>Beautiful, wonderful sights to see And wonderful sounds to hear: The world is the place for a seeing eye And a place for a listening ear.</p> <p>Puppies and lambs and kittens to touch! Satins and silks to feel! Sugar and salt and honey to taste! Fragrant fruits to peel!</p> <p>Beautiful, wonderful, pleasant world! And a child who would know it well Has everything to see and hear, To touch and taste and smell.</p> <p style="text-align: right;">James S. Tippet</p>	<p>C. The poet thinks that the world is-</p> <p>1) sad 2) cold 3) wonderful 4) scary</p> <p>D. The poet thinks the world is a place for-</p> <p>1) The mind 2) Animals 3) The senses 4) Eating</p> <p>E. The poet describes fruit as-</p> <p>1) Tasty 2) Sour 3) Sweet smelling 4) Unpleasant</p>	<p>A. The poem is intended to-</p> <p>1) Give information 2) Give a description 3) Explain how the world appears to the blind 4) Tell how animals feel</p> <p>B. The sights are described as-</p> <p>1) tasty 2) helpful 3) wonderful 4) noisy</p>	<p>A ☐</p> <p>B ☐</p> <p>C ☐</p> <p>D ☐</p> <p>E ☐</p> <p style="text-align: right;">X ☐</p>
---	---	--	--

were transferred into the computer, with the help of an assistant who read the answers aloud, and were scored using the marking program. The program used to score the tests was checked for accuracy against a random selection of twelve hand-scored tests, three from each of the two grade five forms and three from each of the two grade four forms. All scores were then automatically transferred into the Systat® statistical program, version 3.1, running on the Macintosh II, to reduce the probability of entry error due to a transfer of data.

Test Environment:

In each of three schools that participated in the study, a room was assigned for testing purposes. All rooms were illuminated by overhead fluorescent tubes. One of the rooms had a window, however, the drapes remained closed during the test sessions. The rooms were equipped with library tables which could hold two computers.

Room Illumination:

The lighting conditions in the test rooms were evaluated using a Hagner Luminance Meter, model S2, with an internal cell for luminance measurements and an external cell for illuminance measurements. The lighting conformed to the standards published by the Canadian School Trustees Association (Trail, 1988). For classrooms with windows, additional lighting to raise the illumination level to 750 lux was recommended. Since natural light sources were excluded for this study, the illumination levels as

TABLE 4.**Illumination Measures for Testing Rooms By School**

Room Location		Illuminance (lux)
School #1	vertical	180
	horizontal	565
	keyboard	350
School #2	vertical	190
	horizontal	580
	keyboard	364

measured, were slightly lower. Illuminance measurements were taken at each of four computer and paper work areas. The median measurements for each location in the two rooms are shown in Table 4.

The luminances of the the paper test pages and the test pages generated by the two computer monitors were also determined before testing was begun. The computer monitors were arranged perpendicular to the overhead luminaires so as to minimize any reflected glare. Six measurements were taken from a viewing distance of 1.5 meters. Measurements taken from both the paper and the computer generated test sheets included areas with and without text (Fig 5). An analysis of expected reading performance at the highest, lowest and median luminances, calculated using the

Figure 5: Luminance Measurements. Circles indicate the locations from which measurements were taken and letters on the figure correspond to letters in the table of luminance measurements.

beautiful, wonderful sight to see
 And wonderful music to hear:
 The world is the place for a sailing ship
 And for the wind to blow.

Supplemental text and follow to finish:
 Before sailing to sea:
 Sailing ship and sailing to sea:
 The great world to sail.

beautiful, wonderful, pleasant world
 The world is the place for a sailing ship
 The world is the place for a sailing ship
 The world is the place for a sailing ship.
 John A. Tappan

A. The poem is intended to -
 1) Give information
 2) Give a description
 3) Explain how the world appears to the blind
 4) Tell how someone felt

B. Explain and describe -
 1) beautiful
 2) to hear
 3) helpful
 4) to hear

C. The poet thinks that the world is -
 1) sad
 2) wonderful
 3) old
 4) empty

D. The poet thinks the world is a place for -
 1) The wind
 2) A sailboat
 3) The ocean
 4) Sailing

E. The poet describes the world -
 1) Truly
 2) Best
 3) Describing
 4) Unpleasant

F. The poet describes the world -
 1) Truly
 2) Best
 3) Describing
 4) Unpleasant

 Test Page Luminance Measurements (cd m^{-2})

Location	Paper	Monochrome Monitor	Color Monitor
A	72	65	76
B	69	60	80
C	66	61	85
D	69	61	85
E	69	61	85
F	57	55	79

algorithm developed by Rea (1986), demonstrated that reading visual performance (RVP) would vary by no more than 1.1% (Appendix 6).

Procedure:

Following determination of reading level, the subjects within each reading level were randomly assigned to either the high resolution color monitor or the monochrome monitor condition. Within each reading level and monitor condition, the subjects were further assigned to either the computer first or paper first condition. Test forms for each condition were balanced within reading levels and within monitor conditions (Fig 1).

Subjects were removed from their classrooms, in groups of four, and taken to the testing room. For subjects assigned to the paper test condition, a manilla folder, labelled with the subject's first name and last initial and containing the test sheets, was placed at the assigned seating position and the subject was directed to that position. Subjects assigned to the computer test condition were directed to the appropriate computer. General information concerning the test purpose and format was given to all subjects and specific instructions on how to complete the test was given to subjects in the paper condition and computer conditions separately. The test was not started until both sets of instructions had been completed and the subjects signified that they understood the procedure for answering the test questions.

2

MicroD

Task Description:

Subjects in the paper condition were told to open their folders to the first page. The reading paragraphs, questions and answers and answer boxes were pointed out and the connection between the answers and the answer boxes was explained. They were then told to first read the paragraphs on each page and then to read the questions following each paragraph. The number corresponding to each answer selection was pointed out and the subjects were told to select the best answer and to put the number which appeared beside that answer in the answer box.

Instructions for the reading speed subtest were given following completion of the comprehension subtest. The subjects were told to read the story, which covered three pages and, to watch for missing words in some of the sentences. For sentences with a missing word, the subjects were told to choose the best word to finish the sentence from a selection of three words. The reading speed subtest was answered in the same manner as the comprehension subtest. The subjects selected the best answer and put the number which appeared in front of the word in the answer box

For subjects in the computer test condition, the same instructions were given except that the subjects were told to use the red keys on the keyboard to put in their answers. The coded keys were pointed out and the subject was shown how to operate the 'Go' key, which was explained by telling the subjects that it permitted

them to 'go on to the next question'. The 'Next Page' key was demonstrated when all subjects were prepared to begin the test session. The subjects in the computer condition were told to press the next page key, which closed and saved the cover sheet of the computer version of the test, demonstrating how each completed page of the test could be 'turned'. The test was then completed using the coded keys to record the answers in the answer boxes.

In both the computer and the paper test condition, all subjects were given twenty minutes to complete the comprehension subtest to avoid the possibility that the faster readers might complete the entire subtest in less than the thirty minute time interval specified for administering the SDRT. The length of the reading speed subtest was such that the specified three minute time limit was maintained. The two versions of the test were administered on consecutive days.

Results

Order, Form, Sex and Grade Effects:

Although test order and test form were balanced within reading level and grade, the ratio of male to female subjects could not be balanced across all reading levels. Also, the use of two different grades in the study raised the possibility that children of different ages may have reacted differently to some of the subtests. In order to examine these possibilities, the scores obtained by the subjects on the paper version of the SDRT were, due to the different length of the grade four comprehension subtest, standardized separately for each of the two grades and then combined. The z-scores were then subjected to a multiple analysis of variance to determine whether grade, sex, test order or test form might need to be treated as covariates in subsequent analyses. Since the primary hypothesis required a measure of the change in comprehension within subjects rather than a diagnostic approach to determining grade level, the procedure used in Study 1 was maintained for determining percent comprehension of questions completed. The results of the analysis of covariance are shown in Table 5.

Scores on the comprehension subtest, including the literal and inferential scores and the number of questions completed did not show any significant effect for test order, test form, grade or sex. Although the difference in scores on the literal portion of the test due to grade was not significant ($p \leq 0.058$), there was an indication

TABLE 5

Analysis of Variance for Effects of Reading Level, Order, Form, Grade and Sex on The Subtests of the Stanford Diagnostic Reading Test

Variable	Covariates	df	F-Ratio
Literal	Level	2	87.917*
	Form	1	<1.0
	Order	1	1.016
	Grade	1	3.734
	Sex	1	< 1.0
Inferential	Level	2	60.330*
	Form	1	< 1.0
	Order	1	< 1.0
	Grade	1	< 1.0
	Sex	1	2.359
Number Completed	Level	2	38.858*
	Form	1	<1.0
	Order	1	2.262
	Grade	1	< 1.0
	Sex	1	3.045
Reading Rate	Level	2	4.745*
	Form	1	1.787
	Order	1	<1.0
	Grade	1	2.724
	Sex	1	2.674
Context Prediction	Level	2	3.440**
	Form	1	< 1.0
	Order	1	< 1.0
	Grade	1	12.153*
	Sex	1	< 1.0

* $p \leq 0.01$

** $p \leq 0.05$

that grade four subjects may differ on literal comprehension scores. An examination of the means for each grade showed that average and above average grade four subjects obtained literal comprehension scores of 95% and 94% respectively, compared to 78% for the grade five subjects. Since the proportion of grade four and five subjects

was the same in each group, and only literal scores differed between grades, this did not cause any problem throughout the analyses. The reading rate subtest, which included the context prediction scores, showed no significant effect for test order, test form or sex. The scores on context prediction did, however, show a significant effect for grade which did not reduce the effect for reading level. This was not unexpected since the context prediction subtest was taken from the grade five version of the SDRT. Rather than use grade as a covariate in subsequent analysis, the two grades were analyzed separately for this test (Table 10).

Reading Level Assignment:

In order to establish the accuracy of teacher assessment of reading level, a comparison of the reading level assessments made by the teachers with the predicted reading level assessment, using the centroids obtained from a discriminant analysis of the scores obtained on the SDRT, was performed. The predicted assignments for subjects in each of the two grades are shown in Table 6.

Table 6

A: Comparison of Teachers Assessments of Reading Skill and Test Assignment of Reading Level for Grade Five Subjects

		Predicted Reading Level			
		1	2	3	TOTAL
Assigned	1	16	0	0	16
Reading	2	1	15	0	16
Level	3	1	0	15	16
TOTAL		18	15	15	48

Table 6**B: Comparison of Teachers Assessments of Reading Skill and Test Assignment of Reading Level For Grade Four Subjects**

		Predicted Reading Level			
		1	2	3	TOTAL
Assigned	1	7	1	0	8
Reading	2	2	6	0	8
Level	3	1	0	7	8
TOTAL		10	7	7	24

There was a 92% agreement between teacher assessment and the prediction from the discriminant analysis for subjects in both grades. As a precaution, further analyses included a measure to detect outliers since up to six of the subjects, depending on the variable to be analyzed, did not match the categories in which they had been placed.

Monitor Resolution:

The two monitors used in the study differed in resolution to the extent that the difference in pixel density was noticeable. Since adult studies have found a decrease in reading speed with lower monitor resolutions, scores obtained by subjects on the two monitors used in this study were examined to determine whether subjects writing the computer version of the SDRT differed in scores due to monitor resolution. The previously standardized

TABLE 7.

Test for Differences Between Monitor Resolutions On Literal, Inferential and Total Comprehension Subtests on the Stanford Diagnostic Reading Test.

Variable	Source	Sum-Of Squares	DF	Mean Square	F-Ratio
Literal	Level	0.324	2	0.162	14.404*
	Resolution	0.023	1	0.023	2.009
	Level*Resolution	0.017	2	0.008	< 1.0
	Error	0.742	66	0.011	
INFERENTIAL	Level	0.473	2	0.237	17.204*
	Resolution	0.000	1	0.000	< 1.0
	Level*Resolution	0.003	2	0.002	
	Error	0.886	66	0.014	
TOTAL	Level	0.473	2	0.237	17.204*
	Resolution	0.006	1	0.006	< 1.0
	Level*Resolution	0.001	2	0.001	< 1.0
	Error	0.621	66	0.009	
READING RATE	Level	0.473	2	0.237	17.204*
	Resolution	299.370	1	299.370	< 1.0
	Level*Resolution	150.702	2	75.351	< 1.0
	Error	522399.582	66	7916.145	
CONTEXT PREDICTION	Level	0.650	2	0.325	11.104*
	Resolution	0.000	1	0.000	< 1.0
	Level*Resolution	0.115	2	0.057	< 1.0
	Error	1.932	66	0.029	

* $p \leq 0.01$

scores from subjects in the two grade levels were subjected to a multivariate analysis of variance to determine whether reading level, resolution or an interaction between resolution and reading level contributed to score differences between subjects on the two computer monitors. The results of the analysis are shown in Table 7.

There was no effect due to resolution on any of the variables included in the study. Nor was there any interaction effect of resolution with reading level. The difference in the resolution of the two monitors, therefore, did not affect the scores obtained by the subjects on the computer version of the test.

Since the paper form of the test was able to discriminate levels and the subjects were accurately placed within categories regardless of monitor condition, the non-significant results of the comparison of monitor resolution permitted subsequent analyses to be performed on groups collapsed across resolutions.

Comparison Of Paper-Computer Comprehension Scores:

Having established that the paper version of the test discriminated between reading levels and that monitor resolution did not contribute to differences between subjects, the influence of reading from computer monitors both between and within subjects was addressed. The standardized scores obtained by subjects on the paper version of the test and the computer version of the test were compared using a multivariate analysis of variance to determine whether reading level contributed to any differences in scores seen between media. The results of the between subjects analyses are shown in Table 8. Significant differences in scores obtained between subjects due to reading level were demonstrated by the analysis, therefore subsequent analyses were performed by reading level.

Table 8

Tests for Differences Between Reading Levels On Literal, Inferential and Total Standardized Comprehension Scores on the Paper and Computer Versions of The Stanford Diagnostic Reading Test Questions.

Variable Paper/Computer	Source	Sum of Squares	DF	Mean Square	F-Ratio
Literal	Level	53.534	1	53.534	100.191*
	Error	36.334	68	0.534	
Inferential	Level	54.679	1	54.679	76.350*
	Error	48.699	68	0.716	
Total	Level	63.829	1	63.829	112.237*
	Error	38.672	68	0.569	

* $p \leq 0.01$

A within subjects analysis of the scores obtained on the paper version of the SDRT and the computer version of the SDRT was performed, by reading level, to determine where differences in scores occurred. The results of the within reading levels analysis are shown in Table 9. The results of the within reading levels analysis demonstrated a reduction in the comprehension subtest scores obtained on computer by readers in the average reading level group. The scores obtained by the average readers on the computer version of the SDRT were more than one standard deviation lower than the scores obtained on the paper version of the SDRT for both the literal and inferential portions of the subtest. The scores obtained by the below average readers on the computer version of

Table 9

**Within Reading Levels Contrast Of Paper - Computer
Standardized Scores on the Comprehension Subtest of the
Stanford Diagnostic Reading Test.¹**

Variable	Reading Level	Contrast	Mean	DF	F-Ratio
LITERAL	Above Average	Paper	0.827	1,23	< 1.0
		Computer	0.831		
	Average	Paper	0.331	1,23	30.580*
		Computer	-0.452		
	Below Average	Paper	-1.021	1,21**	11.356*
		Computer	-0.366		
INFERENTIAL	Above Average	Paper	0.738	1,23	1.143
		Computer	0.841		
	Average	Paper	0.360	1,23	19.553*
		Computer	-0.420		
	Below Average	Paper	-1.042	1,21**	9.352*
		Computer	-0.461		
TOTAL	Above Average	Paper	0.822	1,23	1.642
		Computer	0.935		
	Average	Paper	0.366	1,23	44.956*
		Computer	-0.472		
	Below Average	Paper	-1.092	1,21**	13.309*
		Computer	-0.480		

* $p \leq 0.01$

* * Outliers removed

the comprehension subtest were two thirds of a standard deviation higher than the scores obtained on the paper version of the subtest. A post hoc analysis of the scores obtained on computer showed that the average and below average reading levels could be discriminated on paper ($p \leq 0.000$, df 1,69) but could not be discriminated on computer ($p \leq 0.849$, df 1,69). Scores for subjects in the above average reading group did not differ between the computer version and the paper version of the subtest for either of the comprehension subtest variables. The interaction between media and reading level is demonstrated in Fig. 6.

A separate analysis of silent reading rate and context prediction scores on the reading rate subtest showed no significant difference in silent reading rate between media for either grade (Appendix 7). An analysis of context prediction scores did, however, result in significant difference between media as shown in Table 10. The scores for above average grade four subjects on the computer version of the context prediction subtest were significantly lower than scores obtained on the paper version of the subtest. Context prediction scores for average and below average readers did not differ significantly between media.

Because the calculation of comprehension scores used for this study was not the same as the scoring procedure used by the SDRT, a comparison of SDRT scores was performed to determine whether actual test scores were also affected. Since raw scores cannot usually be compared across subjects, Stanford scaled scores were calculated and subjected to an analysis of variance. The scaled scores do not take individual reading speed into account, therefore,

Figure6: (a) Mean Standardized Literal Comprehension Scores Obtained on the Paper and Computer Modified Versions of the Comprehension Subtest of the Stanford Diagnostic Reading Test. Mean Scores represent the ratio of questions correct to questions completed.

(b) Mean Standardized Inferential Comprehension Scores Obtained on the Paper and Computer Modified Versions of the Comprehension Subtest of the Stanford Diagnostic Reading Test. Mean Scores represent the ratio of questions correct to questions completed.

An asterisk identifies mean scores on the literal and inferential components of the computer version of the SDRT which differ significantly from mean scores obtained on the literal and inferential components of paper version of the SDRT (n = 24 [above average and average]; 22 [below average] $p \leq 0.001$).

A.

B.

TABLE 10.

**Within Reading Levels Contrasts Of Paper-Computer
Standardized Scores on the Context Prediction Subtest of the
Stanford Diagnostic Reading Test, by Grade.**

Reading Level	Grade	Contrast	Mean	DF	F-Ratio
Above	4	Paper	0.927	1,6**	6.570*
		Computer	0.369		
Average	5	Paper	0.732	1,15	< 1.0
		Computer	0.648		
Average	4	Paper	-0.289	1,7	1.272
		Computer	0.209		
	5	Paper	-0.249	1,15	< 1.0
		Computer	0.041		
Below	4	Paper	-0.522	1,7	< 1.0
		Computer	-0.531		
	5	Paper	-0.447	1,14	< 1.0
		Computer	-0.645		

* $p \leq 0.05$

** Outlier removed

some of the subjects, who were assigned to reading levels based solely on the teachers assessment of level of comprehension, appeared as outliers in the analysis of scaled scores (Appendix14). Four above average slow readers and three below average fast readers were classified as outliers, however, the distribution did not differ from the distribution obtained with ratio scores (Fig. 7).

Figure 7: Stanford Diagnostic Reading Test Scaled Scores Obtained on the Paper and Computer Modified Versions of the Comprehension Subtest of the Stanford Diagnostic Reading Test.

An asterisk identifies scores on computer version of the SDRT which differ significantly from mean obtained on the literal and inferential components of paper version of the SDRT (n = 20 [above average]; 24 average]; 22 [below average] $p \leq 0.05$).

Part 2: Response Device

To assess the possibility that the use of a keyboard to input answers to the computer version of the test may have contributed to any observed changes in comprehension scores, it was necessary to include a condition in which subjects used the same method to answer questions on both forms of the test. The use of a pencil and paper to answer questions presented on a computer monitor was rejected due to the extra visual-motor activity required in moving from the computer monitor to the surface of a table. Keyboard input for both paper and computer versions of the test appeared to be a more suitable approach to input standardization. The results for the comparison group are reported separately to facilitate comparisons between the scores for the keyboard input condition and the pencil condition.

Because the number of available grade four subjects was higher than that of grade five subjects, and because of the small number of below average readers available for study in the participating schools, a decision was made to initially restrict the subjects used in this study to above average and average grade four subjects. If the results indicated a need for a wider sample, grade five subjects would be added at a later date.

The study duplicated the design of Study 2, using the same repeated measures approach. Average and above average readers in grade four completed both the comprehension subtest and the silent reading speed subtest on computer and on paper using a keyboard to

input answers to the questions on both versions of the test. The instructions given to the subjects were identical to the instructions given to subjects who used a pencil to complete the paper version of the test with the exception that instructions for use of the keyboard were given during both test sessions.

Method

The design of the main study was followed to ensure that test forms and test order were balanced within reading levels. For the paper version of the test, a computer version of the same form of the test was loaded to the desktop of one of the computers and the subject was provided with the keyboard as a question answering device. Answers to the paper version of the test, therefore, were entered directly into a computer test folder.

Subjects:

Eight above average and eight average grade four readers were randomly selected from the subjects who had returned forms signifying consent to participate in the study.

Procedure:

Since both computers were required for this study the subjects were taken in pairs to the testing room. They were then asked to sit at their assigned places and the procedure for the study was explained. One of the keyboards was moved from its usual position to provide an answering device for the subject assigned to the paper version of the test. The subject was allowed to move the

keyboard to a comfortable position which permitted an uninterrupted view of both the numeric keypad and the test sheets. The second subject was placed in front of the remaining computer. During the test session, the experimenter sat in front of the monitor of the computer used to record the answers to the paper version of the test to ensure that the subject recalled the instructions for turning pages.

Task Description:

The subjects were given the same instructions as subjects who participated in the main study. The question sheets were explained and both subjects received instructions concerning the use of the numeric keypad. After ensuring that both subjects understood the instructions, the test session was begun.

Results

To determine whether the presence of a keyboard as a question answering device may have contributed to the changes in comprehension scores observed in Studies 1 and 2, the non-standardized scores obtained on paper by the grade four subjects who participated in Study 2 were compared to the scores obtained on paper by the grade four subjects who used the keyboard in this study. The results of the comparison are presented in Table 11.

There was no effect for input mechanism on any of the variables analyzed. Nor was there an interaction between input and reading level. The presence or absence of the keyboard, therefore, did

TABLE 11

Test for Differences Between Keyboard and Pencil as an Input Device On Literal, Inferential and Total Comprehension and Reading Rate and Context Prediction Subtests on the Stanford Diagnostic Reading Test.

Variable	Source	Sum Of Squares	df	Mean Square	F-Ratio
LITERAL	LEVEL	0.039	1	0.039	< 1.0
	INPUT	0.836	1	0.836	< 1.0
	LEVEL* INPUT	0.118	1	0.118	< 1.0
	ERROR	28.977	27	1.073	
INFERENTIAL	LEVEL	3.933	1	3.933	4.475*
	INPUT	1.054	1	1.054	1.200
	LEVEL* INPUT	1.205	1	1.205	1.371
	ERROR	23.728	27	0.879	
TOTAL	LEVEL	2.542	1	2.542	2.580
	INPUT	0.213	1	0.213	< 1.0
	LEVEL* INPUT	0.644	1	0.644	< 1.0
	ERROR	26.605	27	0.985	
READING RATE	LEVEL	1.967	1	1.967	2.001
	INPUT	0.033	1	0.033	< 1.0
	LEVEL* INPUT	1.331	1	1.311	1.353
	ERROR	26.550	27	0.983	
CONTEXT PREDICTION	LEVEL	0.404	1	0.404	25.467**
	INPUT	0.022	1	0.022	1.357
	LEVEL* INPUT	0.000	1	0.000	< 1.0
	ERROR	0.428	27	0.016	

* $p \leq 0.05$

** $p \leq 0.01$

not affect the scores obtained by the subjects on the two versions of the SDRT. The two reading levels could be discriminated on paper on the basis of the scores on inferential comprehension although, as

noted in the main study, scores on literal comprehension for grade four subjects were high for both groups and could not be discriminated.

The keyboard did not interfere with the ability of the paper version of the SDRT to discriminate reading levels on inferential comprehension scores in grade four, nor did it affect the scores on the average and above average readers using a keyboard compared to similar subjects using a pencil on the same subtests. The effect of the keyboard on differences in scores obtained on the computer version of the comprehension subtest of the SDRT compared to the paper version of the SDRT, however, also required examination. Therefore, a within levels analysis of the scores obtained by subjects using the keyboard to answer questions on the paper version of the test compared to the scores obtained by the same subjects on the computer version of the test were performed. The results of the analyses are summarized in Table 12.

As found in both previous studies, scores obtained by subjects assessed as above average did not change between the two media on either the literal or inferential portions of the comprehension subtest. The scores obtained by subjects assessed as average, however, were lower on the computer version of the comprehension subtest than those obtained on the paper version of the comprehension subtest.

The reading rate and context prediction subtests, taken from the grade five version of the SDRT, was analyzed separately. The results of the analysis are shown in Table 13.

Table 12

Within Reading Levels Contrast Of Paper Keyboard-Computer Scores on the Comprehension Subtest of the Stanford Diagnostic Reading Test.

Variable	Reading Level	Contrast	Mean	df	F-Ratio
LITERAL	Above Average	Paper	0.961	1,6	< 1.0
		Computer	0.953		
	Average	Paper	0.970	1,7	10.342*
		Computer	0.920		
INFERENTIAL	Above Average	Paper	0.893	1,6	< 1.0
		Computer	0.883		
	Average	Paper	0.804	1,7	8.264*
		Computer	0.734		
TOTAL	Above Average	Paper	0.927	1,6	< 1.0
		Computer	0.918		
	Average	Paper	0.886	1,7	32.037*
		Computer	0.829		

* $p \leq 0.01$

** $p \leq 0.05$

Table 13

Within Reading Levels Contrast Of Paper Keyboard-Computer Scores on the Reading Rate and Context Prediction Subtests of the Stanford Diagnostic Reading Test.

Variable	Reading Level	Contrast	Mean	df	F-Ratio
READING RATE	Above Average	Paper	276.57	1,6	< 1.0
		Computer	255.86		
	Average	Paper	267.50	1,7	< 1.0
		Computer	256.75		
CONTEXT PREDICTION	Above Average	Paper	0.895	1,6	17.529*
		Computer	0.696		
	Average	Paper	0.662	1,7	< 1.0
		Computer	0.696		

* $p \leq 0.01$

As previously shown (Appendix 6), the scores on the reading rate subtest did not change for subjects in either reading level, however, the scores on the context prediction portion of this subtest showed a significant decrease in the scores obtained by above average readers on the computer version of the subtest. The average readers demonstrated a small increase in the scores obtained on the computer version of the context prediction subtest, however, the change was not significant.

Summary

The results obtained using a balanced design, an automated answering format and the complete comprehension and silent reading speed subtests of the SDRT paralleled the results of the preliminary study. Subjects assessed as above average obtained similar scores on both a paper version and a computer version of the SDRT. Subjects assessed as average obtained higher scores on a paper version of the SDRT than they did on a computer version of the same subtest although there was no indication that their reading speed was significantly affected when the presentation medium was a computer rather than paper. Subjects assessed as below average also did not demonstrate a reduction in reading speed with computer as a presentation medium, however, the scores obtained on the computer version of the SDRT comprehension subtest were significantly higher than scores obtained on the paper version of the same subtest. In fact the scores of the below average readers increased sufficiently that, unlike the paper version, the computer version of the SDRT could not discriminate between average and below average readers. Both changes in comprehension scores and the inability to distinguish the average and below average reading levels were replicated when scaled reading performance scores were used as the dependant variable. Individual reading speeds within reading levels, however, influenced reading level assignments when scaled scores were used. Fast readers with low levels of comprehension and slow readers with high levels of

comprehension were classified as outliers based on their scaled scores on the paper version of the test.

Since the findings of the preliminary study were replicated using high resolution monitors that virtually reproduced a sheet of paper, the possibility that monitor resolution or the reversed image might have been responsible for the observed changes in comprehension scores was eliminated. The possibility that a motor component might have affected the scores was also eliminated since the use of a keyboard did not interfere with the scores obtained on paper by grade four subjects, either between reading levels or within media conditions.

The context prediction portion of the reading rate subtest did not show any changes in scores across media, except in the case of the above average grade four subjects who demonstrated a decline in scores on the computer version of the test. Although there were only seven subjects in the above average group in Study 2, the results were replicated in the keyboard comparison study. Because comprehension of the material was a factor in the context prediction subtest, this finding in grade four subjects was followed up in the previously planned text difficulty study.

Chapter 4

STUDY 3

Study 1 and Study 2 demonstrated that comprehension scores were different on computer compared to paper for average and below average readers. The differences in comprehension scores were replicated on monitors with three different resolutions and were not affected by the use of a keyboard on the paper version of the test. Since the ability to read fluently is affected by the difficulty of the text compared to the reading level of the subjects, and since average readers show lower scores on computer while below average subjects show higher scores on computer, a study was devised to determine whether this result could be manipulated by varying the difficulty level of the test material.

Two approaches to the study were possible; administering the grade five test to grade four readers, or administering the grade four test to grade five subjects. Since the grade five context prediction test had already been administered to the grade four subjects and the grade five comprehension subtest may have been too difficult for the average grade four reader, lowering the test difficulty for the average and below average grade five subjects was selected. Above average grade five subjects were not included in this study because of the low difficulty of the test material. To ensure an accurate assessment of the subject's reading levels, subjects for this study were drawn from those subjects who had previously participated in the preliminary study.

Part 1: Reduced Difficulty Reading Material

Method

Subjects

Seven average and seven below average grade five readers were selected from the subjects who had participated in the preliminary study. Pilot study subjects were not entirely naive, however, the test format and test conditions in the pilot study were dissimilar enough that this was not considered to be of major concern.

Procedure:

The subjects were taken from their classrooms to the testing room and were given the same instructions as those given to the subjects in the main study. The procedure for testing used in the main study was also maintained with the exception that subjects were allowed only fifteen minutes to complete the grade four version of the SDRT since the test was substantially shorter. The issue of the grade level of the test material was not raised with the subjects nor was the grade level indicated anywhere on the test.

Results

The grade four version of the test was able to distinguish between the two groups on the paper version of the test (Fig. 8), therefore the test was not below the reading ability of the average readers. To determine whether the groups differed on the scores obtained on the paper version of the test compared to the computer version of the test, a within levels analysis of variance was

Table 14

Within Reading Levels Contrasts Of Paper - Computer Scores For Grade Five Subjects on the Grade Four Test of the Stanford Diagnostic Reading Test.

Variable	Reading Level	Contrast	Mean	DF	F-Ratio
TOTAL	Average	Paper	0.883	1,6	< 1.0
		Computer	0.900		
COMPREHENSION					
SCORE	Below Average	Paper	0.784	1,5**	9.845*
		Computer	0.706		

* $p \leq 0.05$

**one subject was not available for the second test session

performed for each of the segments of the comprehension subtest. The results of the analyses are shown in Table 14.

Subjects assessed as average did not differ on the scores obtained on the two versions of the test (paper mean; 88.3%, computer mean 90.0%; $p \leq 0.617$). Subjects assessed as below average obtained lower total scores on the computer version of the comprehension subtest than on the paper version (paper mean; 78.4%, computer mean 70.6%; $p \leq 0.026$). Figure 8 shows the results of the reduced text difficulty on comprehension scores. Neither the literal nor inferential score differences achieved significance individually, however, since only 5 below average subjects were available for the second test session, the probability of obtaining significance on each comparison was reduced.

Figure 8: Mean Standardized Scores Obtained by Grade 5 Subjects on the Paper and Computer Modified Versions of the Comprehension Subtest of the grade 4 Stanford Diagnostic Reading Test. Mean Scores represent the ratio of questions correct to questions completed. An asterisk identifies mean scores on the literal and inferential components of the computer version of the SDRT which differ significantly from mean scores obtained on the literal and inferential components of paper version of the SDRT (n= 7 [average]; 6 [below average] $p \leq 0.05$)

To determine the extent to which reading material difficulty, as compared to reading level, contributed to the changes in scores between the media, a multiple analysis of variance was performed on the scores obtained on paper and those obtained on computer by the grade five subjects who wrote the grade four test and the grade five subjects in Study 2 who wrote the grade five test. The analysis of variance tested whether comprehension score changes were attributable to reading level, test medium or material difficulty, or whether a combination of variables contributed to the change in comprehension scores.

The results of the analysis, summarized in Table 15, showed that both reading level ($p \leq 0.01$) and reading material difficulty ($p \leq 0.01$) contributed to score differences, however, when reading level was removed, the medium on which the test was written did not contribute significantly to the changes in scores ($p \geq 0.10$), nor did it interact with material difficulty ($p \geq 0.10$). Test medium interacted significantly with reading level alone ($p \leq 0.05$), however, there was a three way interaction between test medium, reading level and material difficulty ($p \leq 0.01$).

Since a test with a lower reading difficulty interacted with reading level and test medium for the below average readers and average readers, further investigation was necessary to determine whether the same interaction occurred for all reading levels. If so, the effect of administering a test of high difficulty to a group of average and above average readers should result in a decrease in

TABLE 15.

Test for Interactions Between Reading Levels, Material Difficulty and Media On Total Comprehension Subtest Standardized Scores on the Stanford Diagnostic Reading Test.

Source	Sum-Of Squares	DF	Mean Square	F-Ratio
Level	20.957	1	20.597	48.514*
Difficulty	3.352	1	3.352	7.759*
Medium	0.024	1	0.024	< 1.0
Level * Difficulty	3.886	1	3.886	8.997*
Medium * Difficulty	0.013	1	0.013	< 1.0
Level * Medium	2.293	1	2.293	5.308**
Level * Medium * Difficulty	4.261	1	4.261	9.864*
Error	35.422	82	0.432	

* $p \leq 0.01$

** $p \leq 0.05$

computer scores for above average readers and an increase in scores for average readers if the difficulty level of the material was responsible for the change in performance between the two media. Because the grade five test was designed to establish reading levels as high as post-high school, attempting to administer this test to average grade four readers was not feasible. The context prediction subtest of the silent reading speed test of the SDRT was, however, assessed as 'easy grade five reading level' and could be used to determine the effects of difficult material on grade four readers. Since there were a large number of grade four subjects who had previously written this subtest and some grade four subjects who were still available for testing, the complete comprehension subtest of the grade four level of the SDRT, with the grade five context prediction subtest, was administered to the remaining available subjects and the scores on the paper version and the

computer version of the context prediction subtest were examined for the entire group.

Part 2: Increased Difficulty Reading Material

Method

Subjects

Twenty-two above average and twenty four average grade four subjects who had either returned a consent form and had not been selected for study, or who had participated in one of the previous studies, were used for this analysis.

Procedure:

The scores on the context prediction portion of the silent reading speed subtest of the SDRT, for all above average and average grade four subjects, were collapsed for analysis. All subjects had written the complete comprehension and silent reading speed subtests as described in the previous studies.

Results

Subjects in the above average reading level group showed a significant reduction in scores obtained on computer, while subjects in the average reading level group showed a significant increase in the scores obtained on computer (Table 16). The distribution of scores resembled that found between scores on the computer and paper version of the comprehension subtest for average and below average subjects (Fig 9).

TABLE 16.

Within Reading Levels Contrasts Of Paper-Computer Context Prediction Standardized Scores for Grade Four Subjects on the Stanford Diagnostic Reading Test.

Variable	Reading Level	Contrast	Mean	DF	F-Ratio
Context Prediction	Above Average	Paper	0.757	1,20**	5.953*
		Computer	0.206		
	Average	Paper	-0.365	1,23	5.867*
		Computer	0.127		

* $p \leq 0.05$

** Outliers removed

Summary

The results obtained by analyzing the combined group of grade four subjects supported the results suggesting that text difficulty contributed to the changes in comprehension scores seen between media. The more difficult reading material affected the above average readers on the computer version of this subtest in a manner similar to the average readers on material that was at grade level. The readers in the average group reacted to the more difficult material by showing an increase in scores on the computer version of the test, similar to that seen in the below average group in Study 2. This finding, however, is based on a combined group of 48 subjects whereas previously, only six to eight subjects were used to achieve a significant result. The format of this particular

Figure 9: Mean Standardized Context Prediction Scores Obtained by grade 4 Subjects on the Paper and Computer Modified Versions of the Reading Speed Subtest of the Stanford Diagnostic Reading Test. Mean Scores represent the ratio of questions correct to questions completed. An asterisk identifies mean scores on the computer version of the SDRT which differ significantly from mean scores obtained on the paper version of the SDRT (n=22 [above average] 24 [average]; $p \leq 0.05$).

subtest may have been responsible for the large number of subjects required, nevertheless, the results are not as convincing as those found with smaller subject groups.

One approach that would support this finding arises from the method of grade level test construction. The comprehension subtest or the SDRT, and other tests that are designed to assess grade level ability, are constructed with increasingly difficult questions. If text difficulty interacts with reading medium and reading ability at all levels, below average readers should do better on the computer version of a test where the questions are slightly above their reading level, average readers should do worse on the computer version of a test that is at their reading level and above average readers should do equally well on either version of the test. This hypothesis could be tested graphically by calculating the difference in the cumulative scores achieved on the paper and computer versions of the comprehension subtest and plotting the resulting difference scores against question number for each of the three groups. Using this approach, higher scores on the computer version of the test would result in a line that was below the X-axis, while higher scores on the paper version of the test would result in a line above the X-axis.

Figure 10 demonstrates the results of this calculation. All scores are based on the sum of correct answers as a percentage of the total questions available on the subtest. For grade five subjects, each correct answer represented one-sixtieth of the test, for grade four subjects, each correct answer represented one-forty-eighth of the test. For each subject, correct scores on the computer version

of the test were subtracted from correct scores on the paper version of the test and the mean of the differences in scores was calculated. Higher scores on the paper version of the test, therefore, resulted in positive difference scores while higher scores on the computer version of the test resulted in negative difference scores. Cumulative difference scores obtained by grade four and five subjects were calculated separately due to the differing lengths of the subtest for the two grades.

The distribution of the differences in cumulative scores suggested that, the difficulty of the reading material did, indeed contribute to the changes seen in scores across media and that text difficulty had an effect on the scores of all three reading levels. For the grade five subjects, scores on the first eleven to thirteen questions were higher on the computer version than on the paper version of the comprehension subtest of the SDRT for the below average subjects. After question thirteen, the benefit conferred by the computer diminished and, by question thirty, neither paper nor computer demonstrated superiority. Few questions, however, were answered after question thirty. For the average subjects, scores for the first thirty questions were higher on the paper version of the comprehension subtest than on the computer version. After question thirty, the benefit conferred by the paper version of the test diminished and the computer version began to demonstrate superiority. Again, however, few questions were answered after question forty-five. Above average subjects did equally well on the

Figure10: (a) Mean Cumulative Difference in Scores Obtained by grade 5 Subjects on the Paper and Computer Modified Versions of the Comprehension Subtest of the Stanford Diagnostic Reading Test.

(b) Mean Cumulative Difference in Scores Obtained by grade 4 Subjects on the Paper and Computer Modified Versions of the Comprehension Subtest of the Stanford Diagnostic Reading Test.

The X-axis represents the question numbers. The Y-axis represents the mean cumulative percent of questions answered correctly on paper subtracted from the mean cumulative percent of questions answered correctly on computer. Cumulative scores below the X-axis represent a higher percentage of correct answers on the computer version of the SDRT. Cumulative scores above the X-axis represent a higher percentage of scores answered correctly on the paper version of the SDRT.

A.

B.

two versions of the test for all except the most difficult questions. Following question forty, there was a sharp increase in questions answered correctly on the paper version of the test and incorrectly on the computer version of the test.

The grade four subjects demonstrated a similar trend, however, below average grade four subjects did not show a diminished response to the computer version of the test until questions twenty-nine to thirty. Average subjects showed a similar shift from paper to computer superiority at the same point in the test. Again, for the majority of the test, the above average subjects did equally well on the two versions of the test. However, as question difficulty increased, cumulative scores for the grade four average and above average subjects became indistinguishable, demonstrating that the above average readers answered the more difficult questions correctly on paper but not on computer. Therefore, while the medium on which reading material is presented is responsible for the changes observed in the comprehension scores in studies 1 and 2, the difficulty of the material presented determines the direction of the change.

FOLLOW-UP STUDIES

STUDY 4

Establishing that the difficulty of the material presented on the computer contributes to the changes in comprehension scores does not establish why there is a difference between presenting text on a computer compared to paper. The adult literature relating to reading speed and accuracy suggested that the presence or absence of the ability to read rapidly and accurately may have contributed to the interaction between reading ability and reading medium. The adult literature suggested that most computer monitors interfered with attentional skills on proof-reading tasks (Gould 1985, 1987). Since below average readers have frequently been identified as lacking in selective attention skills (Allington and Strange, 1977) a study was designed to determine whether the computer monitor selectively affected attentional skills, thus improving or interfering with the comprehension ability of average and below average readers.

Method

To establish whether selective visual attention contributed to the observed changes in comprehension scores, a task similar to that used in the adult research reports was developed. Two reading paragraphs, without the questions, were taken from form A and form B of the 1974 version of the Stanford Diagnostic Reading Test. Paragraphs were selected from the test difficulty range at which the greatest separation of paper and computer scores were observed

on the cumulative scores for both the average and below average subjects (Appendix 11). The paragraphs were entered into the Excel® program and words were chosen into which letter reversals were introduced. Words selected for letter reversals were restricted to those in which letter reversals would not produce a combination of letters that resembled a word. Letter reversals that produced a different, but acceptable, word were also excluded.

The paper version of the test was produced by printing the completed pages on a laser printer. The development of instructions, selection of appropriate letter reversals and calculation of a reasonable time limit for completion of the task was accomplished with the assistance of grade four and five volunteers from neighborhood schools.

Subjects:

Twelve average and twelve below average grade five subjects, who had participated in either Study 2 or Study 3 were selected for participation in this study.

Procedure:

Subjects were taken in groups of four to the testing room and assigned to either the paper or computer testing condition. The purpose of the task was explained and instructions were given as to how to complete the task. When the subjects signified that they understood the procedure, the test was begun. Two and one half minutes were allowed for completion of the task.

Task Description:

The subjects were asked to scan two pages of reading material for the purpose of identifying errors in words. The subjects were told that letter reversals may or may not have been present in each line and that there may have been none, one, two or three errors in any one line. Identification of errors was recorded by placing a '1' in the box at the end of each line for each reversal identified and a '0' if no reversals were present (Fig 11). The subjects were told that it was necessary only to find the errors and that they did not have to read the material for content.

Results

Test scores were obtained for both omission and commission errors and for the number of errors scanned in two and one half minutes. The scores were first analyzed for differences between the two reading levels on the paper version of the test. The results, as shown in Table 17, were significantly different between reading levels for omission errors, but not for commission errors. Analyses of variance were then performed, by reading level, contrasting the scores obtained on computer and the scores obtained on paper. The results of the analyses are summarized in Table 18.

TABLE 17.

Tests for Differences Between Reading Levels On Omission, Commission and Error Scanning Rate on the Paper Version of the Selective Attention Task.

Variable	Source	Sum Of Squares	df	Mean Square	F-Ratio
Omission Errors	Level	0.021	1	0.021	5.214*
	Error	0.088	22	0.004	
Commission Errors	Level	0.013	1	0.013	<1.0
	Error	0.423	22	0.019	
Error Scanning Rate	Level	2.042	1	2.042	< 1.0
	Error	1113.917	22	50.633	

* $p \leq 0.05$

Neither omission nor commission scores for the below average subjects differed between media. The number of errors scanned by the below average readers on the computer version of the test, however, was significantly lower than the number of errors scanned on the paper version of the test.

For the average subjects, the number of omission errors doubled on the computer version of the test. The number of errors scanned, however, was the same on both versions of the test. Commission errors did not change between media.

Figure11: Sample Test Page Layout for the Selective Attention Test.
Reading paragraphs for this test were taken from an earlier version of the grade 5 level of the SDRT. NOTE - The test page reproduced here does not contain material found in the Stanford Diagnostic Reading Test and is intended to illustrate only the presentation Format of the test.

During the tornaod that struck Woodleigh last >
 Marhc, many ojbects not meant for flying became >
 jet-propelled. What a peculiar sight it was to see >
 Mrs Brown skimming overhead wiht her umbrella >
 still in her hadn! Luckily, she landde safe adn sonuc' >
 when hre wind catcher deposited her inside her own >
 front yard. The sign fro Mr. Jones' drugstore did not >
 fare as well. The Road inot town is nwo decorated >
 with a rather lareg notice that informs visitors that >
 we porvide "Seventeen flavors of ice-cream!" >

...

..... >

...

..... >

...

..... >

...

..... >

TABLE 18

Within Reading Levels Contrast Of Paper-Computer Scores For the Selective Attention Test

Variable	Reading Level	Contrast	Mean	DF	F-Ratio
Omission Errors	Average	Paper	0.080	1,11	10.040*
		Computer	0.181		
	Below Average	Paper	0.140	1,11	< 1.0
		Computer	0.135		
Commission Errors	Average	Paper	0.125	1,11	< 1.0
		Computer	0.114		
	Below Average	Paper	0.172	1,11	< 1.0
		Computer	0.186		
Error Scanning Rate	Average	Paper	20.917	1,11	1.105
		Computer	18.583		
	Below Average	Paper	21.500	1,11	5.935**
		Computer	16.167		

* $p \leq 0.01$

** $p \leq 0.05$

Summary

The results of this study indicate that the changes in comprehension scores seen between media in the previous studies may be partly due to changes in selective attention attributable to computer monitors. In Studies 1 and 2, average readers showed a decline in comprehension scores on the computer version of the

SDRT. In Study 3, with reading material that was of a lower difficulty level, average readers did not show a decline in comprehension scores. On the selective attention test, average readers demonstrated an increase in omission errors. Scanning rate on the selective attention test did not change between media, nor did silent reading rate change between media for average readers in Study 2. The selective attention test, therefore, suggests that average readers may find attending to detail on computer monitors more difficult than attending to detail on paper, however, this difficulty does not lead to a reduction in reading speed.

The subjects in the below average group demonstrated an increase in comprehension scores on the computer version of the SDRT. In the selective attention test, there was no change in omission errors between media, suggesting that the below average group did not find attending to detail on computer monitors different than attending to detail on paper. The error scanning rate for the below average group did change between media, however. Below average readers demonstrated a significant decline in error scanning rate on computer over paper, unlike the silent reading speed scores seen in Study 2, which did not change between media. The decline in scanning rate did not result in an improvement in error detection rate, possibly due to the fact that error detection relies on spelling skills which below average readers frequently lack (Brown and Smiley, 1977)

Research reports pertaining to adult readers suggest that a visual interference effect may reduce reading rate from computer

monitors. The reduction in scanning rate seen in the below average readers may indicate that the computer monitor forces the below average reader to use a slower reading speed when performing a task which requires sustained vigilance. Since a reduction in reading speed was not detected in the silent reading speed test, nor in an analysis of the number of questions completed on the two versions of the test, one hypothesis that might explain this result is that the below average reader attends more closely to material on a computer monitor which, in turn, enables comprehension questions to be answered from assimilated information rather than from a search of the material. If the reader were no longer required to search the material for answers, the reduction in scanning speed would be offset by an increase in the rate at which questions could be answered. This does not explain, however, why oral reading rate is unaffected by the medium on which material is presented (Table 1).

For average readers, failure to reduce their scanning rate when reading from the computer monitor may have resulted in an inability to detect reading errors in the computer version of the test. Usually, a fluent reader can detect and correct words that have been misread if the errors interfere with comprehension of the text (Allington and Strange, 1977). The average readers experienced more difficulty detecting errors on the computer version of the vigilance task but there was no reduction in their scanning rate. This finding suggests that it might be necessary to exercise increased vigilance when reading from a computer monitor, which probably would result in a slower reading rate for most readers. If this is the case, the

increased errors of omission exhibited by the average readers on the vigilance task may partly explain the decrease in scores obtained by the average readers on the computer version of the comprehension subtest.

Such an explanation can only be tentative, however, since the text used in this study was intended to be scanned for errors and did not have to be read for comprehension. Additionally, based on the results reported in the adult literature, monitors with resolutions as high as those used in this study should have resulted in little, if any, change in scanning rates or accuracy levels across media. The results do, however, suggest that visual attention is different on computer than on paper, and that the effect varies with reading level or, alternatively, with the difficulty of the material given the subjects' reading level. For difficult material, it appears that the computer might cause the reader to use a slower scanning rate whereas with material that is not perceived as difficult, scanning rate is not affected. However, in the absence of a study which includes eye-tracking, any effect of the computer on saccadic movements cannot be determined.

Computer Familiarity:

One factor that might have contributed to the ability to read from a computer monitor is familiarity with computers. Subjects who own and regularly used a computer might be expected to develop, through practice, an increased ability to sustain vigilance on a computer monitor. If access to computers varied between the

different reading level groups, the differences in the scores obtained by the three groups might be partly explained by this.

Throughout the study, subjects were asked whether their families owned a computer or, if not, whether they were familiar with using a computer. Approximately fifty percent of subjects in all three groups had access to a computer at home, therefore pursuing the question of ownership did not seem to be a useful exercise. Access to computers in the schools did differ, however. Although all the schools who participated in the study had computers to which the subjects had access, one of the schools had a large number of computers available for use by the students and an established computer usage program which provided the students with several hours of computer time each week. The comprehension subtest scores were re-analyzed to determine whether differences in computer familiarity between schools interacted with reading level to contribute to the difference between computer and paper test version scores. Table 19 summarizes the results of the analysis.

The analysis of variance did not demonstrate any differences in scores between subjects due to schools, therefore it does not appear that the presence or absence of a computer program affects the differences in scores seen between the paper and computer versions of the comprehension subtest. Since there was no

TABLE 19

Test for the Effect of Computer Familiarity On Differences Between Paper and Computer Literal, Inferential and Total Comprehension Scores on the Stanford Diagnostic Reading Test.

Variable	Source	Sum Of Squares	df	Mean Square	F-Ratio
LITERAL	LEVEL	0.564	2	0.282	29.446*
	SCHOOL	0.003	2	0.002	< 1.0
	LEVEL * SCHOOL	0.047	4	0.012	1.229
	ERROR	0.604	63	0.010	
INFERENTIAL	LEVEL	0.661	2	0.330	24.320*
	SCHOOL	0.005	2	0.003	< 1.0
	LEVEL * SCHOOL	0.042	4	0.011	< 1.0
	ERROR	0.856	63	0.014	
TOTAL	LEVEL	0.606	2	0.303	43.832*
	SCHOOL	0.000	2	0.000	< 1.0
	LEVEL * SCHOOL	0.042	4	0.010	1.504
	ERROR	0.435	63	0.007	
NUMBER COMPLETED	LEVEL	256.367	2	128.183	1.417
	SCHOOL	57.146	2	28.573	< 1.0
	LEVEL * SCHOOL	68.300	4	17.075	< 1.0
	ERROR	5697.589	63	90.438	

* $p \leq 0.01$

interaction between schools and reading level, computer familiarity does not seem to explain the interaction between reading level, text difficulty and reading medium. Without a study comparing computer-naive and computer-familiar subjects, a difficult task given the extent of computer use in the schools, no firm conclusions can be made concerning the effects of familiarity with computers on comprehension scores.

CHAPTER 5

DISCUSSION

The microcomputer offers new approaches to research that combine convenience, accuracy and analytic power in a portable form. The novelty of the medium, however, also means that there is no established educational research design to test for any effects of using the computer as an alternative reading medium. Much of the research reported on the use of computers by children has proceeded on the assumption that reading from a computer monitor is not different than reading from a text-book or a sheet of paper but that the novelty of the medium provides an advantage when teaching children who respond poorly to classroom instruction. The methodologies used in past research have produced results supporting this assumption. Careful consideration of previous research, however, reveals inconsistencies that may have resulted from flawed methodologies. The use of subjects from only one reading level, failure to directly compare reading from the different media and inadequate control of the visual environment may all have contributed to incorrect conclusions.

Experimental Design

The methodology used in this study to examine the effect on comprehension of presenting reading material on a computer monitor included controls over many of the environmental and experimental factors not considered in previous studies. The rooms in which the test sessions were conducted were similar in terms of illumination and care was taken to eliminate any sources of disability glare or

veiling glare. Three different monitors were used to present the reading material; one of which resembled the monitors currently found in schools while the other two could reproduce a virtual page image. The luminances of the virtual page monitors and the paper test sheets were closely matched and, since the monitors produced virtual page images, the screen image closely reproduced the appearance of the test sheet. According to the analysis method of Rea (1986), visual factors contributed, at most, to a reduction in reading visual performance of 1.1% (Appendix 6).

The repeated measures design of this study permitted a direct comparison between comprehension scores obtained on paper and comprehension scores obtained on computer. The Stanford Diagnostic Reading Test, with separately standardized subtests of comprehension and reading speed, provided two equivalent forms of the test and a choice of difficulty levels. The equivalent forms meant that subjects could write a similar test on computer and on paper on consecutive days, which reduced the possibility of maturation effects inherent in any study of children that requires an extended period between testing. Equivalent forms also allowed a direct comparison of the standardized scores obtained by the subjects on a computer and a paper version of the test.

Since a keyboard would be used for the computer version of the test while a pencil would be used for the paper version of the test, a group of subjects was included in Study 2 to examine whether using a keyboard as a response device affected comprehension scores. The procedure followed by these subjects was identical to that followed

by the subjects in the main study with the exception that a keyboard was used to answer questions on both the computer and paper versions of the test. The use of the control group permitted an analysis to be performed to determine whether the keyboard contributed to any score changes observed.

Unlike previous studies, this study included three groups of subjects at different reading levels. The subjects selected were in grade four and grade five where the basic skills required for reading would have been acquired but classroom reading groups would still be used, making teacher assessment of reading skills feasible. Initial assignment to reading level group was provided by teacher assessments of the reading ability of each subject. The accuracy of the teachers assessments was close to 95% as determined by a discriminant analysis of assigned reading levels compared to predicted reading levels calculated from SDRT scores (Table 3; Table 6). The pre-planned use of three different reading level groups permitted analyses to be performed both within and between each of the three groups. Any interactions that occurred within different reading levels could, therefore, be detected.

Major Findings

The results of both Study 1, which used a computer monitor with a dark background and green text, and Study 2, which used monitors with a light background and black text, demonstrated that a computer monitor is not equivalent to paper in terms of reading comprehension. On a between reading level groups comparison, the scores obtained by all subjects on the paper version of the SDRT

clearly distinguished the three different reading levels, as expected. In contrast, the scores obtained on the computer version of the SDRT separated only two different reading levels (Appendix 7).

In Study 1, analyses of individual reading levels revealed a 14% decrease in the scores obtained on the computer version compared to the paper version of the comprehension subtest by subjects in the average reading level group. In Study 2, a decrease of 10% was observed in the computer scores compared to paper scores for the average reading level group. Scores for the below average readers were also different on computer compared to paper versions of the comprehension subtest. In Study 1, comprehension scores increased by 18% on the computer version compared to the paper version of the comprehension subtest, while in Study 2, there was an increase of 13% for the subjects in the below average reading group. The length of the test used in Study 1 was insufficient to draw any conclusion for the above average subjects, however in Study 2, there was no difference between the scores of the above average readers on the paper version and the computer version of the comprehension subtest.

The changes in comprehension scores were replicated in two separate studies using three different monitor resolutions. Although the score changes appear to be more extreme in Study 1 than in Study 2, a comparison of resolutions in Study 2 revealed that there was no effect of resolution on comprehension in that study (Table 7). The monitors used in Study 2 did not differ greatly in terms of resolution or vertical scan rate, however, they both

differed substantially from the monitor used in Study 1. Because of the difference in test length and test format, a direct comparison between Study 1 and Study 2 is not possible. The larger score changes in Study 1 may be attributable to the method of selecting reading paragraphs for the shorter version of the test, or to the differences between the monitors used. However, since the test is designed to be of increasing difficulty, the procedure used to construct the short version of the test increased the steepness of the rate of change in difficulty and possibly increased the differences in comprehension scores between the two media.

Although the results from studies with adults suggest that differences between reading from paper and reading from a computer reside primarily in the speed at which a person can read from the two media (Gould, et.al., 1987), there were no significant differences in either oral reading rate in Study 1 (Table 1) or silent reading rate in Study 2 for any of the reading level groups (Appendix 7). An overall analysis between subjects in Study 2 revealed that there was a tendency to read more slowly from computer compared to paper on the silent reading rate subtest, however, when analyzed by grade, the tendency was found to be restricted to subjects in grade four (Appendix 8). Since the silent reading rate subtest was intended for grade five subjects, this indicates that there might be a subtle effect of the monitor on reading rate related to the difficulty of the material. This finding requires further exploration to determine whether the difficulty of the reading material affects reading speed for text presented on a computer.

The importance of reading speed to overall reading comprehension has not been established (Salvia & Ysseldyke, 1985). Because the study was not designed to provide diagnostic data, both Study 1 and Study 2 excluded the effects of reading speed between subjects on reading comprehension scores by using a correct-to-complete score ratio rather than normative scores based on the SDRT scoring procedure. Although there was no significant difference between the number of questions completed on the two media, the subtle effect of reading speed seen in the overall analysis suggested that an analysis of scores obtained using a standard scoring procedure was warranted. Two approaches were, therefore, selected; a non-parametric analysis of raw scores (Appendix 11) and an analysis of variance of the SDRT scaled scores obtained on computer and on paper (Fig. 7). Both analyses demonstrated changes in comprehension scores similar to those found with the ratio scores used in Studies 1 and 2. Average subjects received lower scores on computer than on paper and below average subjects obtained higher scores on computer than on paper.

The results of Study 1 and Study 2 clearly indicated that presenting text on computers affects reading comprehension. Study 2 showed that the results were not due to differences in image quality, contrast or luminance, since both paper and computer were matched for these factors. Contrary to previous research, this study demonstrates that the effect of presenting reading material on a computer monitor depends primarily on the reading level of the subject and not on the quality of the equipment since the pattern of

comprehension scores for paper and computer were the same for low and high resolution monitors.

The results of these two studies are not, however, entirely at variance with previous research. In studies of adult readers, Gould (1987) observed that the speed at which an adult could read from a computer monitor varied with the quality of the screen display. When reading material was presented a low resolution monitor, reading speed and accuracy was reduced. When high resolution monitors were used, reading speed and accuracy was not affected. The results of Study 2, in which both monitors had good quality images, supported this finding. With the exception of grade four readers, subjects in Study 2 showed no difference in silent reading speed from computer compared to paper. The difference in resolution between the two monitors did not affect silent reading speed.

In studies with children, Atkinson (1974) and Grocke (1987) reported that the computer improved the reading scores of reading delayed children. Candy and Edmonds (1982) implied that there was a difference between the response of average achievers and below average achievers to a computer presented language arts remedial program. The scores obtained by the below average readers in Study 1 and Study 2 support the findings of Atkinson (1974) and Grocke (1987). In both studies, below average readers obtained higher scores on computer than on paper. The scores obtained by the average readers explain the observations reported by Candy and Edmonds (1982). Average readers obtained lower scores on

computer than on paper in both Study 1 and Study 2 which may explain why the average achieving subjects in the Candy and Edmonds study did not benefit from the computer-based language arts tutorial.

The response of the above average readers appears, at least initially, to support the finding of no difference between media reported by Hepfner (1985) and asserted by Hawkins (1984) and Tucker (1983). However, unless all subjects in these studies were above average readers, the results obtained by the above average readers in this study cannot be compared to the results reported in the previous research. A more feasible explanation for the non-significant results obtained in earlier studies is that the analysis was unable to detect any differences. Support for this explanation can be found in Study 2. When the scores obtained by all subjects, regardless of reading level, are subjected to a repeated analysis of variance, there was no significant difference between media (Appendix 9). Selecting subjects without regard to reading level, therefore, masks the effect of the reading medium. Had this study not been designed to overcome the methodological flaws of the previous research, the influence of reading medium on different reading levels would not have been detected.

Reading Level and Material Difficulty Interaction

The possibility of finding an interaction between reading level and reading medium was suggested by the contradictory results reported in the literature. Studies in which only one reading level

was examined, usually employing below average readers, reported an increase in reading performance when the computer was used (Grocke, 1983; Riel, 1983), whereas studies which ignored reading level reported no difference in reading performance on the computer (Hepfner, et. al. 1985). Because these studies did not include direct comparisons between media, however, there was no indication as to what might be responsible for the interaction between reading level and reading medium. The methodology used in this study was capable of detecting an interaction between the reading medium and reading level and the results of Study 1 and Study 2 demonstrated that such an interaction did occur. It did not provide an explanation as to why the changes in scores were different for different reading levels.

Because the media were matched for image quality, contrast and luminance in Study 2, and since visual factors could not account for the ten percent difference in comprehension (Appendix 6), the changes in comprehension scores might have been due to factors specific to the subject groups rather than to factors specific to the reading medium. Since the subject groups were composed of children of average intelligence who differed in reading ability as assessed by their teachers, and Blaxall and Willows (1984) reported that the difficulty of the reading material interacted with reading performance, one method of examining the interaction was to manipulate the difficulty of the test material. Blaxall and Willows' (1984) found that when the material read by below average readers was appropriate to their reading level rather than their grade level, the proportion of errors made by the below average readers did not differ from the errors made by average readers. This indicates that

the skills required for fluent reading may not be absent in below average readers but that the below average readers are not be as skilled in the flexibility required for more difficult reading as are more fluent readers. Based on these findings, average readers may be average only in relation to the material they are asked to read, therefore, if below average readers were to read material at their reading level, they should read and comprehend as well as an average reader. If the reading level-task difficulty interaction also occurred on computer, then logically, reducing the difficulty of the test material should cause the below-average subjects to behave on computer in a manner similar to that of the average readers in Study 2. If this occurred, the interaction would then be a reading level-medium-material difficulty interaction, not simply a reading level-medium interaction. Although this would not explain why the interaction occurred, a positive result would indicate that reading processes contributed to the interaction.

To test this hypothesis, the procedure used in Study 2 was replicated in grade five readers using the grade four level of the SDRT. The results of the study supported the hypothesis that the interaction observed was between reading level, the difficulty of the reading material and the medium on which the material was presented. When the difficulty of the test material was below grade level, that is closer to the ability level of the below average reader, the scores of the below average readers were lower on computer than on paper (Fig 8). The scores obtained by the average readers in

Study 3 resembled those obtained by the above average readers in Study 2, that is they were the same on both versions of the test.

The context prediction portion of the silent reading speed subtest made it possible to extend this analysis to grade four subjects. Although the grade five comprehension test was beyond the skills of the average grade four readers, the SDRT manual classified the text in the silent reading speed subtest as 'easy' grade five reading material. The material should, therefore, have been slightly difficult for the average grade four subject and slightly easy for the above average grade four subjects. Since this subtest required subjects to read the text for comprehension in addition to speed, and since the subtest was used for all subjects in Study 2, an analysis was possible on the effect of increasing the difficulty of the test material on comprehension. If an interaction did occur between reading medium, text difficulty and reading level, when the difficulty of the test material was increased, above average grade four readers should have obtained lower scores on the computer than on paper and average grade four readers should have obtained higher scores on the computer than on paper.

An analysis of all available grade four average and above average subjects supported this hypothesis. On easy grade five material, above average grade four subjects obtained lower scores on computer than on paper while average subjects obtained higher scores on computer than on paper (Fig 9). The larger variability of the scores obtained by subjects on this test necessitated an increased number of subjects relative to the other studies (total

n=48) which may reduce the robustness of the interaction for this subtest. The test was not a pure comprehension test, however, and this may be partly responsible for the greater variability in scores.

Graphing the cumulative distribution of the difference between scores obtained by each of the groups on the computer and paper versions of the SDRT in Study 2 supported the findings from the test difficulty study. The below average readers answered more questions correctly on computer than on paper during the first half of the test. As the questions became more difficult, however, the advantage of the computer diminished and few questions were answered correctly in either medium. Average readers showed the opposite effect, with more questions answered correctly on paper during the first half of the test and more answered correctly on computer as the questions became more difficult. Above average readers did equally well on the two media, except at the very end of the test when an advantage for paper was evident.

These findings suggest that the results obtained in Study 1 and Study 2 represent an interaction between the difficulty of the material in relation to the reading level of a subject and the medium on which the reading material was presented. Average subjects, therefore, will show a decrease in comprehension only if the material presented on computer is at their reading level. Similarly, below average readers will show a comprehension increase only if the material presented on computer is above their reading level. For above average readers, the similarity between the computer and

paper comprehension scores would not occur if the difficulty of the reading material were raised to meet or exceed their reading level.

Significance of the Findings

The nature of the interaction revealed by this study creates a problem for computer psychometry. Manipulating the grade level of the paper version of the SDRT presented to the subjects increased or decreased the subject's comprehension scores relative to the difficulty of the material, however, it did not change the linear nature of the scores relative to one another. Average subjects always achieved higher scores than below average subjects, regardless of the difficulty of the test. When the material was presented on computer, not only did the subjects comprehension scores change, the position of the scores relative to one another was also changed, resulting in an inability to discriminate between average and below average subjects (Fig 5). When the material presented on computer was difficult for the subjects reading level, comprehension scores increased to the extent that average and below average subjects could not be distinguished. When the material presented on computer was close to the subjects reading level, comprehension scores decreased, resulting in a decline in comprehension scores for below average grade five readers on the grade four test (Fig. 6). Material that was easy for the subjects did not result in a change in scores, meaning that average grade five readers did equally well on both the paper and computer versions of the grade four test (Fig 6).

The interaction found on the computer version of the SDRT was not, therefore, the same interaction previously seen on paper between reading level and material difficulty. On paper, presenting reading material above a child's reading level resulted in a decrease in comprehension scores, while on computer, similar material resulted in an increase in comprehension scores. An interaction of this type means that paper-based reading tasks, such as tests of reading comprehension and research which use reading as a basis, cannot readily be transferred to the computer.

Research Directly Comparing Media

Previous studies conducted with children have not compared data gathered from both paper and computer conditions in subjects at different reading levels. It is difficult, therefore, to compare the results of this study to those reported previously. Studies such as those conducted by Atkinson (1974), Grocke (1983) and Candy and Edmonds (1982) suggest that the computer influences reading comprehension, however, the designs employed do not permit any comparison of results to be made. Of the studies that have been conducted, the one most closely related to this research was conducted by Hepfner et.al. (1985), in adult readers.

The subjects who participated in the Hepfner et.al. study were not selected on the basis of reading ability. They were, however, assigned to groups based on their reading and television viewing habits and their familiarity with computers. The results for raw scores within groups did not reveal any consistent pattern related to

reading habits, although heavy readers obtained higher print than computer scores. Fast readers, slow readers and light readers did not differ significantly across media. Similarly, the pattern of scores for computer users and computer nonusers did not show a consistent trend. Both computer users and nonusers, as well as videogame users, obtained higher print scores, while video games nonusers did not differ significantly in scores. None of the groups in the Hepfner et. al. study obtained higher scores on the computer version of the test and the results suggested that subjects who regularly used computers showed greater differences between computer and print scores than any of the other subgroups.

On a retrospective questionnaire, subjects in Hepfner et.al. study expressed the opinion that reading from print was faster than reading from a computer monitor. The study was therefore, reanalysed using a ratio of correct to complete questions to determine whether reading speed may have affected the scores. The analysis showed no difference between scores calculated as a ratio, leading the authors to conclude that comprehension of the material was not affected. Unfortunately, the analysis of the ratio of correct answers to completed answers was performed on the total sample of subjects, rather than on the original subgroups. There is no report of the differences in ratio scores obtained by any of the subgroups, nor the difference in ratio scores obtained by subjects with high versus low print scores. In fact, only median scores are reported for each group and no score at all for the total group. From the data provided, it is not possible to determine whether a change in

comprehension scores might have been observed for subjects such as light readers.

Because of the design used in the Hepfner et.al. study, it is not certain whether an analysis of subgroups may have resulted in a pattern of scores similar to those reported here. Although the study used a paper-computer comparison, subjects were required to switch from paper to computer within the same test session and little information is offered on the quality of the screen image, the degree of automaticity of the test presentation or the effect of having to split reading paragraphs across two screens. Additionally, subjects in their study read the paragraphs from a computer monitor but answered the questions on paper, not on a computer. The comparison, therefore, was not entirely between computer and paper. There also appears to be an implicit assumption in their study that television viewing and the time spent reading are related to ability to comprehend. An examination of print scores may have provided more information on comprehension ability.

Although the study is not entirely relevant since it addresses comprehension in adult subjects, there are some points of interest in relation to the present study. Both this study and the Hepfner et.al. study found a difference by group in comprehension scores. Both studies also addressed the question of comprehension using a test designed for the purpose rather than a task assumed to address comprehension. The most significant difference between the two studies was that their study assumed equivalent reading ability for the analysis of comprehension score ratios, while this study

specifically addressed the issue of reading level prior to testing. Had they separately analyzed the ratio scores obtained by subjects in each of the specified categories, the results may have been similar to those of the present research.

Selective Attention and Reading from Computer Monitors

Demonstrating that an interaction occurs between reading level, reading medium and material difficulty explains how comprehension scores differ between media, but does not address the question of why such an interaction should occur. To address this issue, a study which examined individual reading skills was required. Given the large number of skills which are currently accepted as contributing to the reading process, it was not possible to conduct an exhaustive study, however, the results of several studies conducted in adults suggest that both reading speed and accuracy can be affected when text is presented on a computer monitor (Bauer and Cavonius, 1980; Radl, 1980, Gould, 1987). Since these studies used proof-reading ability as an experimental task, and since proof-reading requires selective attention to detail, a study was designed to examine the ability of grade five children to attend selectively to letter reversals within words. Although scanning tasks are not directly related to tasks requiring comprehension, selective attention to detail does appear to contribute to reading ability and below average readers frequently exhibit deficits in this area (Brown & Smiley, 1977).

Three dependant variables were analyzed in the selective attention task; the number of letter reversals identified, the number of letter reversals missed and the number of errors scanned. The results of the study were inconclusive, however, they did suggest that selective attention is affected by the medium on which the reading material is presented. Average readers missed more spelling errors on the computer monitor than they did on paper, however their error scanning speed did not change nor did the number of commission errors. The scores of the below average readers did not change on either omission or commission errors, however, their rate of error scanning was lower on the computer monitor than on paper (Table 18).

Any conclusions drawn from this study can only be tentative. The task selected required a degree of spelling proficiency which may have been lacking in the below average readers, however, the findings did indicate some future research directions. The reduction in error scanning rate demonstrated by the below average readers suggests that, when the material to be read is difficult for the reader, the computer monitor requires a reduction in the rate at which the material is scanned. For average readers, the increase in omission errors, coupled with no change in scanning rate may indicate that a decreased scanning rate will occur only when the material is perceptibly difficult. Since neither average nor below average readers showed a decrease in the number of comprehension questions completed in any of the previous studies, nor any change in silent reading speed rate (Appendix 7), the effect of scanning rate is

probably subtle, especially on a high resolution monitor. This effect does seem to occur for difficult material, however, as can be seen from the overall decrease in silent reading rate observed in the grade four, but not the grade five, subjects.

The results of the selective attention study do not appear to agree with the findings of Bagnara (1980), who reported that scanning accuracy for spelling errors was not reduced in adults when the resolution of a computer monitor was sufficiently high. Since Bagnara also reported that scanning accuracy for repeated words on a high resolution monitor is lower than on paper, however, the findings may not be entirely at variance. The discrepancy in the results may be due to the difference between the subjects used by Bagnara and the subjects used in this study. Bagnara used adult computer operators who were adept at proof-reading tasks and who would be aware that the two tasks, detection of errors in words and detecting repeated words, would necessitate different strategies. This skill is not well developed in children. Although average grade 8 readers are able to adjust their reading speed for different reading requirements, increasing the difficulty of the reading material interferes with this skill (DiStefano, Noe, & Valencia, 1981). Since the material used in this study was not adjusted for reading level, the below average readers may not have been capable of determining an appropriate reading strategy for error detection. Additionally, because of the spelling ability requirement, this task may not have been suitable as a method of measuring attention in below average readers. To determine the impact of selective attention on

comprehension from computer monitors, therefore, future studies should include a different selective attention task.

Recommendations for Further Research

The studies reported here apply only to children in grades four and five. Since children in this age group have usually acquired the basic decoding skills necessary to proceed to reading for comprehension, the effect of reading from a computer monitor on the comprehension of children in the early stages of reading cannot be determined. If below average readers can read grade level material better from a computer than from paper, early reading may be enhanced by the use of a computer monitor. An extensive study, using a wider age range and children at varying skill levels, should be conducted to replicate these findings and to determine whether the computer similarly affects children of different ages.

To determine which skill or process is responsible for the changes seen in comprehension scores, a study is needed in which children with known skill deficits are tested for reading ability on both paper and computer. A wide age range of children would be required to control for maturation effects since many early deficits are age-related. Ideally, such a study would include a comparison of elementary, junior high and high school subjects, matched with adults who have reading deficits which were identified during their school career. This would indicate whether extensive experience with reading can overcome the effects of presenting reading material on a computer monitor. Since saccadic eye movements may

be affected by text difficulty (Ryan, 1981), and since this study did not include measurements of saccades, future studies should include a means of tracking saccadic eye movement.

In addition to studies aimed at identifying the process responsible for the changes in comprehension scores, a study is required in which the fundamental difference between paper and a reversed image monitor is controlled. Since a monitor with a light background is self-luminous, there is a degree of flicker which is not present on paper. To determine whether this difference is a factor in reading from a computer monitor, a comparison between reading from surface illuminated paper, paper lit from behind with a light which produces similar flicker, and text presented on a computer monitor with a light background should be conducted.

Summary and Conclusions

Firm conclusions as to which skill or skills are responsible for the interaction between reading level, test medium and material difficulty level are not possible from this study. However, the results do lead to conclusions regarding current computer programs in the school system. Unlike adults, children are not subject to merely reduced reading speed on high resolution monitors. When reading material at their reading level, children in both grade four and grade five do not comprehend material presented on computer as well as they do when the same material is presented on paper. Material that is slightly above their reading level is comprehended better on computer than on paper, while material that is below a child's reading level can be comprehended equally well on either medium.

These findings suggest that the current approach to computer use in the schools requires careful consideration. A decrease in comprehension of ten to thirteen percent, in a child who relies on reading to acquire information necessary to succeed in school, should not be ignored. In order to overcome the decrease in comprehension, material presented on the computer must be tailored to be either slightly above or slightly below the reading level of the child.

The implications for psychometric testing of children on computer are extremely important. The test used for this study, the Stanford Diagnostic Reading Test, does not provide a valid

assessment of reading level for children reading at or below reading grade level' when administered on computer. The scores obtained by children reading at grade level are lower on the computer version of the test than on the paper version. Scores obtained by children reading below grade level are higher on the computer version of the test than the paper version. The SDRT may not be the only test so affected. Any test which relies on an increasing level of question difficulty to determine reading ability could also be rendered invalid if transferred to the computer. Because of the interaction between the difficulty of the reading material and the reading medium, there does not appear to be any solution to this dilemma.

The implications for students exposed to computers as an alternative reading medium could be significant. For an average reader in grade five, a ten percent reduction in comprehension scores on the complete comprehension subtest would translate into a reduction of one grade level equivalent based on the recommended scoring procedure. For a below average reader, an increase of ten-to-twelve percent would result in a score one grade equivalent higher than actual reading ability. While the average reader might be erroneously assigned to material below his reading ability, the implications for the below average reader could well be that he or she would be denied the assistance needed to improve a skill that is basic to school success. No educational programming recommendations can be based on the results of a computerized test that requires reading unless the test can be shown to be valid on a computer.

The process which is responsible for the change in comprehension between the two media cannot be determined from the studies conducted. The reading ability of a subject, relative to the difficulty of the reading material, determines whether comprehension will be reduced or enhanced by presenting the material on a computer monitor. Reading speed, as a function of text difficulty, appears to be related to the change in comprehension scores but, because of the subtlety of the change in reading speed, may not be a causal factor. Determining the process responsible for the changes in comprehension will require an extensive examination of the reading skills currently accepted as being integral to reading fluency and their relationship to reading from different media. If the skills responsible for the change in comprehension scores between media can be determined from future studies, it may be possible to establish the relative contribution of each skill to fluent reading.

BIBLIOGRAPHY

- Abernathy, A. MacUser Labs: True Colors. MacUser.1988, Vol 4(10), pp 236-271.
- Allington, R. & Strange, M. Effects of Grapheme Substitution in Connected Text Upon Reading Behaviours. Visible Language, 1977, 11, 285-297.
- Atkinson, R. Teaching Children to Read Using a Computer. American Psychologist, 1974, 169-178.
- Backman, J., Mamen, M., & Ferguson, H.B. The Reading Level Design: Conceptual and Methodological Issues in Reading Research. Psychological Bulletin, 1983.
- Bagnara, S. Error Detection at Visual Display Units. In E. Grandjean (Ed.), Ergonomic Aspects of Visual Display Terminals. London: Taylor & Francis Ltd., 1980, 143-146.
- Bauer, D. & Cavonius, C. Improving the legibility of Visual Display Units Through Contrast Reversal. In E. Grandjean (Ed.) Ergonomic Aspects of Visual Display Terminals. London: Taylor & Francis Ltd., 1980, 137-142.
- Blaxall, J. & Willows, D. M. Reading Ability and Text Difficulty as Influences on Second Grader's Oral Reading Errors. Journal of Educational Psychology, 1984, 76 (1), 330-341.

- Bouma, H. Visual Reading Processes and the Quality of Text Displays. In E. Grandjean (Ed.) Ergonomic Aspects of Visual Display Terminals. London: Taylor & Francis Ltd., 1980, 101-114.
- Brown, A. & Smiley, S. Rating the Importance of Structural Units of Prose Passages: A Problem of Metacognitive Development. Child Development, 1977, 48, 1-9.
- Bryant, P. & Bradley, L. Why Children Sometimes Write Words Which They Do Not Read. In U. Frith (Ed.), Linguistic Information and Spelling. London: Churchill Press, 1979, 355-372.
- Candy, L. & Edmonds, E. A Study in the Use of a Computer as an Aid to Teaching English. International Journal of Man-Machine Studies. 1982, 16, 333-339.
- Carlson, N. Physiology of Behavior. Boston: Allyn & Bacon Inc., 1981.
- Carrillo, L. The Language-Experience Approach to the Teaching of Reading. In J. Downing and A. Brown (Eds.), The Second International Reading Symposium. London: Cassell & Co. Ltd., 1967, 197-205.
- Chandler, H. Computers in the Schools. Journal of Reading Disabilities, 1983, 16 (1), 54-56.
- Colombo, E., Kirschbaum, C. & Riatelli, M. Legibility of texts: The influence of blur. Lighting Research Technology. 1987, 19, 61-71.
- Das, J. Simultaneous and Successive Processing in Children With Reading Disability. Topics in Language Disorders, 1984, 6, 34-47.

DiStefano, P., Noe, M., & Valencia, S. Measurement of the Effects of Purpose and Passage Difficulty on Reading Flexibility. Journal of Educational Psychology, 1981, 73 (4), 602-606.

Doehring, D., Trites, R., Patel, P., & Fiedorowicz, C. Reading Disabilities: Perspectives in Neurolinguistics, Neuropsychology and Psycholinguistics. London: Academic Press Inc., 1981.

Dunn, L. Exceptional Children in the Schools: Special Education in Transition. New York: Holt, Rinehart & Winston, Inc, 1973.

Goodman, K. On the Psycholinguistic Method of Teaching Reading. Elementary School Journal, 1971, 177-181.

Goodman, K. The Psycholinguistic Nature of the Reading Process. In K. Goodman (Ed.) The Psycholinguistic Nature of the Reading Process. Indiana: Wayne State University Press, 1976, 15-26.

Goodman, K. What's Whole in Whole Language. Toronto: Scholastic-Tab Publications, 1985.

Gould, J. & Grischkowsky, N. Does Visual Angle of a Line of Characters Affect Reading Speed. Human Factors, 1986, 28, 165-173.

Gould, J., Alfaro, L., Finn, R., Haupt, B., Minuto, A. & Salaun, J. Why Reading Was Slower From CRT Displays Than From Paper. CHI+GI, 1987, 7-11.

Grocke, M. Computers In The Classroom: How Can They Teach Reading. Australian Journal of Reading, 1983, 6 (4), 175-185.

Haslerud, G. & Clark, R. On the Redintegrative Perception of Words. American Journal of Psychology, 1957, 70, 97-101.

Hawkins, J. Microcomputers in Schools: Impact on the Social Life of Elementary Classrooms. Journal of Applied Developmental Psychology, 1982, 3, 361-373.

Heiman, J. & Ross, A. Saccadic Eye Movements and Reading Difficulty. Journal of Abnormal Child Psychology, 1974, 2, 53-71.

Hepfner, F., Anderson, J., Farstrup, A. & Weideman, N. Reading on a standardized test is better from print than from computer display. Journal of Reading, 1985, 321-325.

Holmes, D. The Independance of Letter, Word and Meaning Identification in Reading. Reading Research Quarterly, 1971, 6, 394-415,

Holmes, G. The Relative Legibility of Black and White Print. Journal of Applied Psychology, 1931, 13(3), 248-251.

Howell, D. Statistical Methods for Psychology; 2nd Ed. Boston: Duxbury Press, 1987.

Kirschmann, A. Cited in; Holmes, G. The Relative Legibility of Black and White Print. Journal of Applied Psychology, 1931, 13(3), 248-251.

Kochnower, J., Richardson, E., DiBenedetto, B. A Comparison Of Phonic Decoding Ability Of Normal And Learning Disabled Children. Journal Of Learning Disabilities, 1983, 16, 348-351.

- Lahey, B., McNees, M. & Brown, C. Modification of Deficits in Reading for Comprehension. Journal of Applied Behavior Analysis, 1973, 6, 475-480.
- Legge, G., Rudin, G. & Luebker, A. Psychophysics of Reading-V. The Role of Contrast in Normal Vision. Vision Research, 1987, 27 (7) 1165-1177
- Mark, L., Shankweiler, D., Liberman, I. & Fowler, C. Phonetic Recoding and Reading Difficulty in Beginning Readers. Memory and Cognition, 1977, 6, 623-629.
- Maynard, R. High-Tech Schooling. Homemakers, 1988, 21(7), 12-26.
- Oborne, D. & Holton, D. Reading from Screen versus Paper. International Journal of Man-Machine Studies, 1988, 28, 1-9.
- Orton, S. Reading, Writing and Speech problems in Children. New York: Norton, 1937.
- Papert, S. Mindstorms: Children Computers and Powerful Ideas. New York: Basic Books Inc., 1980.
- Patterson, D. & Tinker, M. Studies of Typographical Factors Influencing Speed of Reading. Journal of Applied Psychology, 1931, 13(3), 241-247,
- Pidgeon, D. Logical Steps In The Process Of Learning To Read. Educational Research, 1976, 18, 174-181.
- Posner, M. & Boies, S., Components of Attention. Psychological Review, 1971, 78, 391-408.

Poynter, H., Sebor, C., Haynes, H. & Hirsch, J. Oculomotor Functions in Reading Disabilities. American Journal of Optometry and Physiological Optics, 1982, 59(2), 116-127.

Radl, G. Experimental Investigations for Optimal Screen Presentation-Mode and Colors of Symbols on the CRT Screen. In E. Grandjean (Ed.), Ergonomic Aspects of Visual Display Terminals. London: Taylor & Francis Ltd., 1980, 127-136.

Rea, M. Overview of Task Performance in the Illuminated Environment. In M.S. Rea (Ed.), Illumination Engineering Learning Aid: Volume 11, 1984.

Rea, M. Practical Implications of a New Visual Performance Model. Lighting Research and Technology, 1986, 18, 113-117.

Rea, M. Toward a model of Visual Performance : A Review of Methodologies. Journal of the Illuminating Engineering Society, 1987, 128-142.

Rea, M. Toward a model of Visual Performance: Foundations and Data. Journal of the Illuminating Engineering Society, 1986, 41-57.

Rea, M. Visual Performance : With Realistic Methods of changing Contrast. Journal of the Illuminating Engineering Society, 1981, 164-177.

Rea, M., Ouellette, M. & Kennedy, M. Lighting and Task Parameters Affecting Posture, Performance and Subjective Ratings. Institute for Research in Construction: paper no. 1361, 1985.

Riel, M. Education and Ecstasy: Computer Chronicles of Students Writing Together. The Quarterly Newsletter of The Laboratory of Comparative Human Cognition, 1983, 5(3), 59-68.

Robinson, P. Macuser Labs: Black and White and Read All Over. MacUser, 1988, Vol. 4(8), pp. 102-144.

Ross, A. Reading and Reading Disorders. In A. Ross (Ed.) Psychological Aspects of Learning Disabilities and Reading Disorders, New York: McGraw Hill, 1976, 141-166.

Ryan, E. Identifying and Remediating Failures in Reading Comprehension: Toward an Instructional Approach for Poor Comprehenders. Reading Research: Advances in Theory and Practice, 1981, 3, 223-261.

Salvia, J. & Ysseldyke, J. Assessment in Special and Remedial Education, Boston: Houghton Mifflin Co., 1985.

Samuels, S. & Edwall, G. The Role of Attention in Reading with Implications for the Learning Disabled Student. Journal of Learning Disabilities, 1981, 14(6), 353-362.

Schwartz, R., & Stanovich, K. Flexibility In The Use of Graphic And Contextual Information By Good And Poor Readers. Journal Of Reading Behavior, 1981, 13, 263-269.

Shankweiler, D. & Liberman, I. Misreading: A Search For Causes. In J. Kavanaugh and F. Mattingly (Eds.) Language By Ear and By Eye. Boston: MIT Press, 1972, 293-317.

Smith, F. Psycholinguistics and Reading. Boston: Rinehart and Winston, Inc., 1973.

Smith, S. & Rea, M. Check Value Verification Under Different Levels of Illumination. Journal of the Illuminating Engineering Society, 1987, 143-149.

Smith, S. & Rea, M. Performance of a Reading Test Under Different Levels of Illumination. Journal of the Illuminating Engineering Society, 1982, 29-33.

Stanovich, K. Individual Differences In The Cognitive Processes Of Reading: I. Word Decoding. Journal Of Learning Disabilities, 1982, 15, 485-493.

Stanovich, K. Individual Differences In The Cognitive Processes Of Reading: II. Text-Level Processes. Journal Of Learning Disabilities, 1982a, 15, 549-554.

Stanovich, K. Toward an Interactive-Compensatory Model of Individual Differences in the Development of Reading Fluency. Reading Research Quarterly, 1980, 16, 32-71.

Swanson, L. Vigilance Deficit In Learning Disabled Children: A Signal Detection Analysis. Journal Of Child Psychology & Psychiatry & Allied Disciplines, 1981, 22, 393-399.

Taylor, C. The Relative Legibility of Black and White Print. Journal of Educational Psychology, 1934, 25(8), 561-578.

Tierney, R. Review of Stanford Diagnostic Reading Test, 1976 Edition. In O. Buros (Ed.) The eighth mental measurements yearbook. New Jersey: The Gryphon Press., 1978, 1178.

Timmers, H., van Nes, F. & Blommaert, F. Visual Work Recognition as a Function of Contrast. In E. Grandjean (Ed.), Ergonomic Aspects of Visual Display Terminals. London: Taylor & Francis Ltd., 1980, 115-120.

Torgeson, J. & Young, K. Priorities For The Use Of Microcomputers With Learning Disabled Children. Journal of Learning Disabilities. 1983, 16 (4), 234-236.

Trail, D. Personal communication, Carleton Board of Education, 1988.

Tucker, M. Computers in Schools. Theory Into Practice. 1983, 22 (4), 313-320.

Wallace, G. & Larsen, S. Educational Assessment of Learning Problems: Testing for Teaching. Boston: Allyn & Bacon, Inc., 1979.

Weston, W. The Relation Between. Illumination and Visual Efficiency-The Effect of Brightness Contrast. In M.S. Rea (Ed.), Illumination Engineering Learning Aid: Volume 11. 1984.

Willows, D. M. & Ryan, E. Differential Utilization of Syntactic and Semantic Information by Skilled and Less Skilled Readers in the Intermediate Grades. Journal of Educational Psychology, 1981, 73 (5), 607-615.

Willows, D. M. & Ryan, E. Development of grammatical sensitivity and its relationship to early reading achievement Reading Research Quarterly, 1981, 73 (5), 607-615.

Willows, D. M. Individual Differences in Distraction by Pictures in a Reading Situation. Journal of Educational Psychology, 1985, Summer, 253-266.

Willows, D. Reading Between the Lines: Selective Attention in Good and Poor Readers. Child Development, 1974, 45, 408-415.

Willows, D., Borwick, D., & Butowski, I. From theory to practice in reading research: Toward the development of better software. In Pressley, M. & Levin, J. (Eds) Cognitive Strategy Research: Educational Applications. New York: Springfield Verlag, 1987, pp 157-187.

Zola, D. Redundancy and Word Perception During Reading. Perception and Psychophysics, 1984, 36 (3), 277-284.

APPENDIX 1

Calculation of Scores

Comprehension score calculation:

	PLC/TLP	PIC/TIP	PL	PI	CLC/TLC	CIC/TIC	CL	CI	LEVEL
1	22/24	19/24	0.917	0.792	18/19	15/19	0.944	0.778	Average
2	23/24	22/24	0.958	0.917	21/21	21/22	1.000	0.950	Average
3	23/24	24/24	0.957	1.000	24/24	22/24	1.000	0.917	Average
4	24/24	20/24	1.000	0.833	13/16	13/16	0.800	0.800	Average
5	15/19	18/20	0.789	0.889	24/24	20/24	1.000	0.833	Average
6	17/19	14/19	0.889	0.722	24/24	21/23	1.000	0.909	Average
7	20/24	21/24	0.833	0.880	22/24	18/24	0.910	0.740	Average
8	21/24	16/24	0.870	0.652	15/19	12/19	0.789	0.632	Below Average
9	20/24	17/24	0.833	0.708	15/21	12/20	0.737	0.600	Below Average
10	22/24	18/24	0.917	0.750	20/24	17/24	0.826	0.696	Below Average
11	13/17	15/18	0.778	0.867	16/21	11/22	0.773	0.524	Below Average
12	21/24	16/24	0.870	0.652	20/22	15/23	0.913	0.636	Below Average
13	21/24	16/24	0.870	0.652	18/24	14/24	0.750	0.583	Below Average
MEAN			0.883	0.793			0.880	0.738	
S. D.			0.066	0.114			0.104	0.139	

z-score calculation:

	(PL-X/S)	(PI-X/S)	(CL-X/S)	(CI-X/S)	LEVEL
1	0.512	-0.012	0.613	0.285	Average
2	1.132	1.085	1.151	1.519	Average
3	1.116	1.813	1.151	1.282	Average
4	1.766	0.348	-0.770	0.443	Average
5	-1.423	0.839	1.151	0.680	Average
6	0.088	-0.626	1.151	1.225	Average
7	-0.758	0.760	0.287	0.012	Average
8	-0.199	-1.241	-0.876	-0.763	Below Average
9	-0.758	-0.749	-1.375	-0.993	Below Average
10	0.512	-0.381	-0.520	-0.304	Below Average
11	-1.590	0.646	-1.029	-1.538	Below Average
12	-0.199	-1.241	0.315	-0.734	Below Average
13	-0.199	-1.241	-1.250	-1.115	Below Average

PLC = Literal questions answered correctly - paper

TLP = Total Literal questions attempted - paper

PL = Literal comprehension scores - paper

PIC = Inferential questions answered correctly - paper

TIP = Total Inferential questions attempted - paper

PI = Inferential comprehension scores - paper

CLC = Literal questions answered correctly - computer

TLC = Total Literal questions attempted - computer

CL = Literal comprehension scores - computer

CIC = Inferential questions answered correctly - computer

TIC = Total Inferential questions attempted - computer

CI = Inferential comprehension scores - computer

Means and Standard Deviations by Grade and Reading Level

STUDY 1

Grade 5

READING LEVEL = ABOVE AVERAGE

TOTAL OBSERVATIONS: 5

	PT	CT	PN	CN	PSPEED	PACC	CSPEED	CACC
NOF CASES	5	5	5	5	5	5	5	5
MEAN	0.879	0.876	24.20	20.800	145.600	0.968	127.800	0.972
S. D.	0.056	0.051	3.271	2.168	14.588	0.008	4.604	0.008

READING LEVEL = AVERAGE

TOTAL OBSERVATIONS: 7

	PT	CT	PN	CN	PSPEED	PACC	CSPEED	CACC
NOF CASES	7	7	7	7	7	7	7	7
MEAN	0.747	0.608	21.714	20.714	146.429	0.939	137.286	0.946
S. D.	0.074	0.166	4.82	4.192	23.557	0.027	25.572	0.016

READING LEVEL = BELOW AVERAGE

TOTAL OBSERVATIONS: 8

	PT	CT	PN	CN	PSPEED	PACC	CSPEED	CACC
NOF CASES	8	8	8	8	8	8	8	8
MEAN	0.470	0.659	17.750	18.625	122.800	0.938	140.875	0.954
S. D.	0.135	0.118	4.334	3.204	12.969	0.026	21.886	0.023

Grade 4

READING LEVEL = AVERAGE

TOTAL OBSERVATIONS: 4

	PT	CT	PN	CN
NOF CASES	4	4	4	4
MEAN	0.921	0.633	19.500	18.000
S. D.	0.054	0.089	2.380	3.559

READING LEVEL = BELOW AVERAGE

TOTAL OBSERVATIONS: 4

	PT	CT	PN	CN
NOF CASES	4	4	4	4
MEAN	0.650	0.817	20.750	16.500
S. D.	0.070	0.042	0.500	3.109

PT= Total comprehension scores - paper

CT= Total comprehension scores - computer

PN= Total number of questions completed - paper

CN= Total number of questions completed- computer

P (&C)SPEED= Oral reading speed (words/min.)

PACC= Oral reading accuracy-paper (words read correctly/ Total words read)

CACC= Oral reading accuracy-computer (words read correctly/ Total words read)

STUDY 2

Grade 5

READING LEVEL = ABOVE AVERAGE

TOTAL OBSERVATIONS: 16

	PL	PI	PT	PN	CL	CI	CT	ON
NOF CASES	16	16	16	16	16	16	16	16
MEAN	0.935	0.855	0.894	45.063	0.908	0.851	0.879	43.625
S. D.	0.051	0.079	0.048	8.354	0.062	0.080	0.056	8.221

READING LEVEL = AVERAGE

TOTAL OBSERVATIONS: 16

	PL	PI	PT	PN	CL	CI	CT	ON
NOF CASES	16	16	16	16	16	16	16	16
MEAN	0.828	0.777	0.802	40.125	0.723	0.634	0.678	40.625
S. D.	0.050	0.088	0.050	10.243	0.093	0.082	0.062	10.462

READING LEVEL = BELOW AVERAGE

TOTAL OBSERVATIONS: 15

	PL	PI	PT	PN	CL	CI	CT	ON
NOF CASES	15	15	15	15	15	15	15	15
MEAN	0.631	0.567	0.597	39.067	0.740	0.701	0.718	34.067
S. D.	0.100	0.126	0.101	10.046	0.105	0.143	0.107	11.190

GRADE 4

READING LEVEL = ABOVE AVERAGE

TOTAL OBSERVATIONS: 8

	PL	PI	PT	PN	CL	CI	CT	ON
NOF CASES	8	8	8	8	8	8	8	8
MEAN	0.950	0.891	0.921	46.125	0.953	0.887	0.920	45.000
S. D.	0.072	0.073	0.065	3.720	0.045	0.047	0.033	3.505

READING LEVEL = AVERAGE

TOTAL OBSERVATIONS: 8

	PL	PI	PT	PN	CL	CI	CT	ON
NOF CASES	8	8	8	8	8	8	8	8
MEAN	0.948	0.866	0.908	37.750	0.881	0.796	0.839	39.750
S. D.	0.039	0.088	0.046	7.649	0.044	0.072	0.049	9.867

READING LEVEL = BELOW AVERAGE

TOTAL OBSERVATIONS: 7

	PL	PI	PT	PN	CL	CI	CT	ON
NOF CASES	7	7	7	7	7	7	7	7
MEAN	0.727	0.552	0.639	37.714	0.872	0.699	0.785	37.571
S. D.	0.087	0.151	0.103	9.742	0.133	0.152	0.124	7.368

PAPER

PL = Literal comprehension score
 PI = Inferential comprehension score
 PT = Total comprehension score
 PN = Total no. of questions completed

COMPUTER

CL = Literal comprehension score
 CI = Inferential comprehension score
 CT = Total comprehension score
 ON = Total no. of questions completed

Grade 5

READING LEVEL = ABOVE AVERAGE

TOTAL OBSERVATIONS: 16

	PR	PP	OR	CP
N OF CASES	16	16	16	16
MEAN	355.375	0.866	296.625	0.836
S. D.	95.666	0.089	62.674	0.136

READING LEVEL = AVERAGE

TOTAL OBSERVATIONS: 16

	PR	PP	OR	CP
N OF CASES	16	16	16	16
MEAN	305.938	0.701	276.375	0.734
S. D.	106.381	0.161	87.759	0.142

READING LEVEL = BELOW AVERAGE

TOTAL OBSERVATIONS: 15

	PR	PP	OR	CP
N OF CASES	15	15	15	15
MEAN	262.467	0.667	257.867	0.621
S. D.	90.346	0.176	87.083	0.151

GRADE 4

READING LEVEL = ABOVE AVERAGE

TOTAL OBSERVATIONS: 8

	PR	PP	OR	CP
N OF CASES	8	8	8	8
MEAN	325.250	0.838	273.125	0.769
S. D.	115.502	0.101	76.333	0.183

READING LEVEL = AVERAGE

TOTAL OBSERVATIONS: 8

	PR	PP	OR	CP
N OF CASES	8	8	8	8
MEAN	232.125	0.613	229.125	0.699
S. D.	120.549	0.162	104.040	0.230

READING LEVEL = BELOW AVERAGE

TOTAL OBSERVATIONS: 7

	PR	PP	OR	CP
N OF CASES	8	8	8	8
MEAN	294.000	0.566	212.875	0.527
S. D.	64.626	0.206	53.914	0.254

PAPER

PR = Silent Reading Rate (words/3min.)

PP = Context prediction score

COMPUTER

OR = Silent Reading Rate (words/3min.)

CP = Context prediction score

STUDY 3

READING LEVEL = AVERAGE

TOTAL OBSERVATIONS: 7

	PL	PI	PT	PN	CL	CI	CT	CN
NOF CASES	7	7	7	7	7	7	7	7
MEAN	0.906	0.862	0.883	45.286	0.951	0.847	0.900	43.286
S.D.	0.075	0.090	0.062	4.645	0.075	0.080	0.070	6.184

READING LEVEL = BELOW AVERAGE

TOTAL OBSERVATIONS: 6

	PL	PI	PT	PN	CL	CI	CT	CN
NOF CASES	6	6	6	6	6	6	6	6
MEAN	0.856	0.714	0.784	45.833	0.798	0.612	0.706	43.833
S.D.	0.047	0.085	0.033	5.307	0.064	0.058	0.053	3.971

PAPER

PL = Literal comprehension score
 PI = Inferential comprehension score
 PT = Total comprehension score
 PN = Total no. of questions completed

COMPUTER

CL = Literal comprehension score
 CI = Inferential comprehension score
 CT = Total comprehension score
 CN = Total no. of questions completed

STUDY 4

READING LEVEL = AVERAGE

TOTAL OBSERVATIONS: 12

	FO	PC	PEE	CO	CC	CEE
NOF CASES	12	12	12	12	12	12
MEAN	0.080	0.125	20.917	0.181	0.114	18.583
S.D.	0.065	0.121	6.999	0.112	0.140	7.798

READING LEVEL = BELOW AVERAGE

TOTAL OBSERVATIONS: 12

	FO	PC	PEE	CO	CC	CEE
NOF CASES	12	12	12	12	12	12
MEAN	0.140	0.172	21.500	0.135	0.186	16.167
S.D.	0.062	0.154	7.230	0.070	0.181	5.686

PAPER

FO = Errors missed
 PC = Mis-identified correct words
 PEE = Total errors encounterable

COMPUTER

CO = Errors missed
 CC = Mis-identified correct words
 CEE = Total errors encounterable

APPENDIX 2

Specifications for Macintosh 11 monitors

	MONOCHROME	COLOR
Dot Per Inch (dpi)*:	76	72
Screen Size (inches)	12	13
Active Screen Area	8.38*6.25	9.5*6.9
Vertical Refresh	67 Hz	65 Hz(approx)**
Horizontal Refresh	35 kHz	35 kHz
Bandwidth	22 MHz	23 MHz
Glare Protection	Etched glass	Dark tint glass

- * A dpi of 72 produces an image that appears identical to a laser-printed page of text. Higher dpi's distort the quality of the image on the screen.
- ** Using an enhanced color card reduces the refresh rate of a color monitor to slightly below the advertised refresh rate. Using an 'aperture grille' (long vertical slits in a wire mask), however, reduces the perceived flicker caused by the reduced refresh rate since slight vertical image shifts do not cause changes in brightness (Mac User Labs, Mac User, October, 1988)

(Source: Apple Technical Manual & MacUser, October, 1988: Macuser Labs)

APPENDIX 3

READING RATE FROM PAPER ORIENTED HORIZONTALLY AND
VERTICALLY (STUDY 1).

Eight grade five children, who participated in Study 1, read a passage orally from the SDRT, presented either vertically or horizontally. After sixty seconds, the subjects were stopped and the last word read was marked. The number of words read in one minute were counted for each condition and a paired samples t-test was performed on the mean words read vertically compared to the mean words read horizontally.

	HORIZONTAL	VERTICAL
N OF CASES	7*	7 *
MINIMUM	91.000	93.000
MAXIMUM	168.000	171.000
MEAN	117.714	116.714
STANDARD DEV	27.066	26.456

PAIRED SAMPLES T-TEST ON HORIZONTAL - VERTICAL ORAL READING RATE
WITH 7 CASES

T = .269 DF = 6 PROB = .797

*One subject was not available for the second test session

APPENDIX 4 CONSENT FORM

Dear Parents,

A project concerned with differences in childrens' ability to read material from computers compared with reading the same material typed on paper is being conducted by researchers from Carleton University. The Research Committee of the Carleton Board of Education has granted permission to the researchers to request your cooperation in this study.

Children who participate in this study will be asked to complete a reading test, once on the computer and once on paper. Total testing time will be about twenty minutes per session and there will be two testing sessions.

Only those children for whom consent has been given by a parent or guardian will be considered for participation in the study. Participation will be based on teacher reports of reading level. The results of any individual child will be held in confidence and no identifying information will be included in the final report. The test used for this study will not affect your child's progress in school nor will the test be used for diagnostic purposes. If you would like your child to participate in this study, please sign the attached form and have your child return it to his/her teacher as soon as possible. If you do not want your child to participate, please indicate and have your child return the form as soon as possible.

I will be pleased to answer any questions you might have concerning this study. I can be reached at 731-9984. Please do not hesitate to call if you require further information

Thank you for your co-operation

Monica D. Dykeman, B.Sc., B.Ed., M.Ed.

PLEASE RETURN THIS PART TO THE CLASSROOM TEACHER

I give my consent for -----
to participate in this study. (NAME OF CHILD: Please print)

I do not give my consent for -----
to participate in this study. (NAME OF CHILD: Please print)

Date

Signature of parent or guardian

APPENDIX 5: Illustration of the keyboard layout for the Computer Version of the Stanford Diagnostic Reading Test. Keys 1,2,3 & 4 on the numeric keypad were used to enter answers. The key labelled 'GO" was used to record answers and automatically advance the cursor. The key labelled 'Next Page" automatically saved each individual test sheet. All keys used by subjects were labelled using red stickers. Remaining keys were non-functional during test sessions.

3 of /de 3

MICROD

APPENDIX 6

Relative Visual Performance calculation using the Algorithm of M. S. Rea. The following information is taken from; Lighting and Research Technology, 18(3), 1986, 113-119.

1. Select task background luminance within the range
 $12 \leq LB \leq 169 \text{ cd m}^{-2}$

2. Calculate threshold contrast C_t at the chosen LB , where

$$C_t = 0.0418 a [(0.308 / bLB)^{0.4} + 1.0]^{2.5}$$

Where a is a factor relating contrast threshold of young observers to that of older observers, having taken into account reductions in retinal illuminance; $a = 1.0$ for twenty-year olds, $a > 1.0$ for older age groups. b is a factor relating retinal illuminance for young observers to that of older observers; $b = 1.0$ for twenty year olds, $b < 1.0$ for older age groups.

3. Select the task contrast C_v where LB is greater than the target luminance L_T and

$$C_v = \frac{LB - L_T}{LB}$$

4. Calculate the parameters n , k and VP_{max} , where

$$n = 0.882 + 4.38 \theta_1 - 6.05 \theta_1^2$$

where $\theta_1 = \log_{10} (\log_{10}(bLB))$.

$$\log_{10} k = -2.25 + 1.77 \theta_2 - 0.217 \theta_2^2$$

where $\theta_2 = \log_{10}(bLB)$

$$VP_{max} = 0.0628 + 0.0120 \theta_2 - 0.00268 \theta_2^2$$

5. Calculate the predicted level of visual performance VP where

$$VP = \{ \Delta C^n / [\Delta C^n + (k / bLB)^n] \} VP_{\max}$$

where $\Delta C = C_v - C_t$. For $\Delta C \leq 0$, $VP = 0$.

6. Visual performance relative to that at $LB = 169 \text{ cd m}^{-2}$ and $C_v = 1.0$ (RVP) may be determined from

$$RVP = VP/f$$

where $f = 0.0760 = VP_{\max}$ at $LB = 169 \text{ cd m}^{-2}$ and $C_v = 1.0$

Using this algorithm, reading visual performance (RVP) was calculated for the highest and lowest background luminances (Fig 5). Since the equipment to determine actual task contrast was not available, maximum contrast was assumed to be not higher than 0.9 and minimum contrast not lower than 0.6. The algorithm does not include factors relating contrast threshold or retinal illuminance for subjects younger than twenty to that of older subjects, however, it was assumed that the contrast threshold and retinal illuminance of twenty-year olds could be applied to ten and eleven- year olds.

For a background luminance of 60 cd m^{-2} , and a task contrast of 0.6

$$\begin{aligned} C_t &= 0.055662 \\ n &= 1.598833 \\ k &= 1.626344 \\ VP_{\max} &= 0.075664 \\ RVP &= 0.987424 \end{aligned}$$

For a background luminance of 85 cd m^{-2} , and a task contrast of 0.9

$$\begin{aligned} C_t &= 0.053724 \\ n &= 1.639285 \\ k &= 2.276433 \\ VP_{\max} &= 0.075976 \\ RVP &= 0.996221 \end{aligned}$$

APPENDIX 7

Within Reading Levels Contrast Of Paper-Computer Scores on the Reading Rate Subtest of the Stanford Diagnostic Reading Test, by Grade.¹ (from Study 2)

Reading Level	Grade	Contrast	Mean	DF	F-Ratio
Above	4	Paper	0.390	1,7	<1
		Computer	0.427		
Average	5	Paper	0.468	1,15	1.143
		Computer	0.161		
Average	4	Paper	-0.486	1,7	1.744
		Computer	-0.114		
	5	Paper	-0.014	1,15	<1
		Computer	-0.072		
Below	4	Paper	-0.079	1,6*	<1
		Computer	-0.219		
Average	5	Paper	-0.486	1,14	2.632
		Computer	-0.059		

¹ Standardized scores

* Outlier removed

APPENDIX 8

Comparison Of Paper - Computer Scores on the Comprehension Subtest of the Stanford Diagnostic Reading Test Without Regard to Reading Level (from Study 2).

Variable	Grade	Contrast	Mean	DF	F-Ratio
LITERAL	5	Paper	0.797	1,47	<1
		Computer	0.788		
	4	Paper	0.864	1,23	1.669
		Computer	0.907		
INFERENTIAL	5	Paper	0.734	1,47	<1
		Computer	0.723		
	4	Paper	0.764	1,23	1.561
		Computer	0.799		
TOTAL	5	Paper	0.765	1,47	<1
		Computer	0.754		
	4	Paper	0.814	1,23	1.812
		Computer	0.853		

APPENDIX 9

Comparison Of Paper - Computer Scores on the Silent Reading Speed Subtest of the Stanford Diagnostic Reading Test Without Regard to Reading Level. (from Study 2)

Variable	Grade	Contrast	Mean	DF	F-Ratio
READING RATE	5	Paper	307.375	1,47	2.630
		Computer	282.646		
	4	Paper	283.792	1,23	5.620
		Computer	238.375		
CONTEXT	5	Paper	0.734	1,47	<1
		Computer	0.727		
PREDICTION	4	Paper	0.672	1,23	<1
		Computer	0.665		

APPENDIX 10

ABSOLUTE SCORES FROM STUDY 2 ON PAPER AND COMPUTER BY GRADE AND READING LEVEL

	GRADE	READING LEVEL	PA	CA	CA-PA
1	Five	Above Average	24.000	46.000	22.000
2	Five	Above Average	27.000	29.000	2.000
3	Five	Above Average	34.000	25.000	-9.000
4	Five	Above Average	36.000	32.000	-4.000
5	Five	Above Average	37.000	42.000	5.000
6	Five	Above Average	38.000	32.000	-6.000
7	Five	Above Average	39.000	39.000	0.000
8	Five	Above Average	42.000	39.000	-3.000
9	Five	Above Average	42.000	45.000	3.000
10	Five	Above Average	43.000	42.000	-1.000
11	Five	Above Average	44.000	36.000	-8.000
12	Five	Above Average	45.000	40.000	-5.000
13	Five	Above Average	46.000	44.000	-2.000
14	Five	Above Average	46.000	49.000	3.000
15	Five	Above Average	49.000	33.000	-16.000
16	Five	Above Average	52.000	39.000	-13.000
17	Four	Above Average	33.000	40.000	7.000
18	Four	Above Average	36.000	41.000	5.000
19	Four	Above Average	41.000	44.000	3.000
20	Four	Above Average	45.000	39.000	-6.000
21	Four	Above Average	45.000	45.000	0.000
22	Four	Above Average	46.000	45.000	-1.000
23	Four	Above Average	47.000	34.000	-13.000
24	Four	Above Average	48.000	44.000	-4.000
25	Five	Average	18.000	16.000	-2.000
26	Five	Average	25.000	33.000	8.000
27	Five	Average	26.000	16.000	-10.000
28	Five	Average	27.000	15.000	-12.000
29	Five	Average	27.000	36.000	9.000
30	Five	Average	28.000	23.000	-5.000
31	Five	Average	28.000	27.000	-1.000
32	Five	Average	28.000	33.000	5.000
33	Five	Average	28.000	20.000	-8.000
34	Five	Average	32.000	28.000	-4.000
35	Five	Average	36.000	20.000	-16.000
36	Five	Average	38.000	34.000	-4.000
37	Five	Average	38.000	34.000	-4.000
38	Five	Average	43.000	33.000	-10.000
39	Five	Average	44.000	30.000	-14.000
40	Five	Average	48.000	45.000	-3.000
41	Four	Average	26.000	39.000	13.000
42	Four	Average	28.000	22.000	-6.000
43	Four	Average	29.000	19.000	-10.000
44	Four	Average	31.000	32.000	1.000
45	Four	Average	34.000	36.000	2.000
46	Four	Average	37.000	36.000	-1.000
47	Four	Average	44.000	40.000	-4.000
48	Four	Average	46.000	42.000	-4.000

**APPENDIX 11:a) Mean Cumulative Scores Obtained by Grade 5
Subjects on the Paper and Computer Modified Versions
of the Comprehension Subtest of the Stanford
Diagnostic Reading Test, by reading level.**

**b) Mean Cumulative Scores Obtained by Grade 4
Subjects on the Paper and Computer Modified Versions
of the Comprehension Subtest of the Stanford
Diagnostic Reading Test, by reading level.**

The X-axis represents the question numbers. The Y-axis represents the mean cumulative percent of questions answered correctly. Cumulative scores were calculated by assigning each question the value of $1/(\text{no. of questions})\%$. [For grade 5 subjects, each question = $1/60\%$; for grade 4 subjects, each question = $1/48\%$]. Correct questions were then summed for each subject and the mean was calculated for each reading group.

**APPENDIX 12: (a) Notched Box Plot of Absolute Number Correct
Obtained by Grade 5 Subjects on the Paper and
Computer Modified Versions of the Comprehension
Subtest of the Stanford Diagnostic Reading Test.**

**(b) Notched Box Plot of Absolute Number Correct
Obtained by Grade 4 Subjects on the Paper and
Computer Modified Versions of the Comprehension
Subtest of the Stanford Diagnostic Reading Test.**

**Above average=reading level 1 (n=16 [grade 5]; n=8
[grade 4]); Average=reading level 2 (n=16 [grade 5];
n=8 [grade 4]); Below average=reading level 3 (n=16
[grade 5]; n=8 [grade 4])**

**Notches represent the median scores for each reading
level. Box ranges represent simultaneous 95%
confidence intervals. Medians for each group that are
not overlapped by the confidence interval of an
adjacent group are significantly different (Howell,
1987).**

A.

B.

APPENDIX 13

MEANS AND STANDARD DEVIATIONS BY TEST FORM

GRADE 5

TOTAL OBSERVATIONS: 24

	PL	PI	PT	PN	CL	CI	CT	ON
N OF CASES	24	24	24	24	24	24	24	24
FORM GB								
Mean	0.806	0.745	0.775	41.875	0.772	0.694	0.732	37.458
S.D.	0.151	0.161	0.151	9.470	0.122	0.171	0.139	9.758
FORM HB								
Mean	0.788	0.723	0.755	40.792	0.803	0.752	0.777	41.625
S.D.	0.142	0.152	0.135	10.043	0.125	0.105	0.097	10.926

	PR	PP	CR	CP
FORM GB				
N OF CASES	24	24	24	24
MEAN	293.375	0.7113	21.875	0.720
S.D.	103.346	0.178	88.258	0.151

	PR	PP	CR	CP
FORM HB				
N OF CASES	24	24	24	24
MEAN	321.375	0.7752	43.417	0.734
S.D.	101.830	0.154	66.670	0.186

PAPER		COMPUTER	
PL	= Literal comprehension score	CL	= Literal comprehension score
PI	= Inferential comprehension score	CI	= Inferential comprehension score
PT	= Total comprehension score	CT	= Total comprehension score
PN	= Total no. of questions completed	ON	= Total no. of questions completed
PR	= Silent Reading Rate (words/3min.)	CR	= Silent Reading Rate (words/3min.)
PP	= Context prediction score	CP	= Context prediction score

MEANS AND STANDARD DEVIATIONS BY TEST FORM

GRADE 4

TOTAL OBSERVATIONS: 12

	PL	PI	PT	PN	CL	CI	CT	CN
NOF CASES	12	12	12	12	12	12	12	12
FORMGG								
Mean	0.858	0.740	0.799	43.333	0.870	0.771	0.821	41.000
S.D.	0.186	0.240	0.209	7.024	0.100	0.138	0.108	8.023

FORMGH								
Mean	0.870	0.789	0.829	38.583	0.945	0.827	0.885	39.167
S.D.	0.105	0.144	0.113	8.618	0.053	0.092	0.058	9.379

	PR	PP	CR	CP
FORMGG				
NOF CASES	12	12	12	12
MEAN	276.000	0.611266	0.083	0.705
S.D.	98.847	0.208	69.753	0.204

	PR	PP	CR	CP
FORMGH				
NOF CASES				
MEAN	291.583	0.734	210.667	0.625
S.D.	117.300	0.171	85.549	0.272

PAPER

PL = Literal comprehension score
 PI = Inferential comprehension score
 PT = Total comprehension score
 PN = Total no. of questions completed
 PR = Silent Reading Rate (words/3min.)
 PP = Context prediction score

COMPUTER

CL = Literal comprehension score
 CI = Inferential comprehension score
 CT = Total comprehension score
 CN = Total no. of questions completed
 CR = Silent Reading Rate (words/3min.)
 CP = Context prediction score

APPENDIX 14

**Within Reading Levels Contrast Of Paper-Computer Scaled
Scores on the Stanford Diagnostic Reading Test.**

Reading Level	Contrast	Mean	DF	F-Ratio
Above Average	Paper	645.5	1,19*	1.392
	Computer	631.3		
Average	Paper	590.2	1,23	4.390*
	Computer	573.9		
Below Average	Paper	538.5	1,21**	5.123*
	Computer	553.9		

* $p \leq 0.05$

** Outliers removed

END

16.02.90

FIN