A Comparison of the Computer Administered and Written Forms of Multiple Choice Testing

by

David Bryan Zandvliet
B.Sc., University of Guelph, 1987

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF ARTS

in the Faculty of Education

ACCEPTED
ULTY OF GRADUATE STUDIES
[redacted]
DEAN
E 15 Oct 93

We accept this thesis as conforming
to the required standard

[redacted]
Dr. P. Farragher, Supervisor (Department of Social and Natural Sciences)

[redacted]
Dr. L. D. Yore, Departmental Member (Department of Social and Natural Sciences)

[redacted]
Dr. J. Anderson, Outside Member (Department of Psychological Foundations)

[redacted]
Dr. L. Devlin, External Member (Department of Communication and Social Foundations)

© DAVID BRYAN ZANDVLIET, 1993

University of Victoria

All rights reserved. Thesis may not be reproduced in whole or in part, by photocopy or other means, without the permission of the author.

ABSTRACT

The design of a computer administered test is described and the potential advantages and disadvantages of this general type of testing is discussed. A computer testing program designed using Hypercard by the author is tested for equivalence with written formats of a test. The study group in this research consisted of fifty adult basic education students enrolled in an introductory computer course. Each student completed a schedule of three written and three computer tests. Additional information about these tests was collected through the use of student pre and post surveys and through the use of an innovative Hypercard script which tracked a student's path through the tests. The results of the comparative analysis of tests scores indicated that there were no significant differences in test scores attributable to the computer or written format of the tests. Additionally, survey responses indicated a student preference for the computer format of the test over the written format and that this preference increased after exposure to the computer administered tests. Finally, an analysis of the student test path data recorded by the computer indicated that the computerized tests took on average two minutes longer for students to complete. The computer test path data and survey responses also gave an indication as to the test design features included in the computer administered test which were most useful to the students as they completed the tests.

[redacted]

Dr. P. Farragher, Supervisor (Department of Social and Natural Sciences)

[redacted]

Dr. L. D. Yore, Departmental Member (Department of Social and Natural Sciences)

[redacted]

Dr. J. Anderson, Outside Member (Department of Psychological Foundations)

[redacted]

Dr. L. Devlin, External Member (Department of Communication and Social Foundations)

Table of Contents

Chapter One 1
Introduction 1
Statement of the Problem 2
Statement of the Sub-problems 2
Hypotheses of the Study 3
Limitations of the Study 3
Importance of the Study 4
Chapter Two 12
Literature Review 12
Chapter Three 36
Research Overview 36
Background of the Study 36
Design of the Program 38
Description of the Pilot Study 42
The Research Study 46
Chapter Four 50
Research Findings 50
Pre -Survey Results 50
Group Demographics 50
Student Responses Prior to Testing 51
Student Responses After Testing 53
Reported Use of Computer Test Features 56
Comparative Analysis of Test Scores 57
Item by Item Analysis of Test Scores 61
Analysis of Test Path Data 64
Comparison of Test Administration Times 64
Frequency of Use of the Test Features 66
Chapter Five 68
Summary and Discussion 68
Summary of Purpose and Methodology 68
Summary of Hypotheses 71
Summary of Findings 72
Summary of Test Administration Times 75
Summary of Test Path Data 76
Conclusions 77
Recommendations from this Research 78
Recommendations for Future Research 79
References 80

List of Tables

Table 1 (Test administration schedule for the study)_____48
Table 2 (Student response summary for pre-survey questions)_____51
Table 3 (Student response summary for post-survey questions)_____54
Table 4 (Overall statistics for scores in test administrations one and two)_____58
Table 5 (Overall statistics for scores in test administrations three to six)_____59
Table 6 (Individual test statistics for test administrations one to six)_____61
Table 7 (Combined test times and statistics for test administrations three to six)__66
Table 8 (Frequency of use of computer test design features)_____67

List of Figures

Figure 1 (Test three incorrect response patterns for different test formats)_____62
Figure 2 (Test four incorrect response patterns for different test formats)_____62
Figure 3 (Test five incorrect response patterns for different test formats)_____63
Figure 4 (Test six incorrect response patterns for different test formats)_____63

Chapter One

Introduction

The use of computers has expanded dramatically during the last decade and this technological growth has led to a corresponding increase in their use within the educational system. The new focus on computer technology has manifested itself not only in an increase in the numbers of computers but also in a diversification of their use. One such use is Computer Administered Testing (CAT). Changes in the educational system have made educators increasingly reliant on tests, and the use of computer technology as a means of assessment has also increased. This brings forward many questions regarding the suitability of the computer as an assessment device (Percival & Ellington, 1984).

Several trends in educational technology are contributing to the increasing focus on computers. The first of these is a gradual shift towards a more student-centered approach in education and naturally a more individualized approach to learning in all its forms. The second trend, an increased appreciation for the motor skills and attitudes of students has led to an increased focus on this type of learning (Province of British Columbia, 1990, Moore & Waldron, 1981). Finally, the almost explosive increase in the use of computers within our larger society has also increased the need to implement computers in our educational institutions.

Statement of the Problem

The administrative benefits of CAT are many. Features such as computerized scoring, evaluation of test items and test editing features can involve considerable time savings for teachers and may allow for increased flexibility in testing, (i.e., more frequent testing and more variability in test scheduling). These factors might eventually allow educators to provide their students with testing on demand. However, despite the convenience of CAT programs, many educators continue to argue that their use may be disadvantageous to students in that students have scored lower on tests given in this format (Wise & Plake, 1990). The problem addressed by this study is the overall suitability of the computer as an assessment device.

Statement of the Sub-problems

For this study the above problem was broken down into three sub-problems. First, the issue of the inequivalency in test results between written and computerized formats was considered with respect to potential cause. Then the potential for increased test anxiety in the computerized format was considered as a possible cause for this inequivalency. Finally, issues related to software and interface designs were related to test anxiety and performance on computerized tests. These areas became the focus for this study. In addition, attempts were made to define the potential advantages and disadvantages of CAT.

Hypotheses of the Study

The three sub-problems of this study were then re-worked into the three hypotheses that would form the basis of this study:

1) The first hypothesis was that a suitable computer administered test could be designed (in consultation with students and teachers) which would elicit a more positive student attitude towards CAT than paper-pencil tests.

2) The second hypothesis was that the computerized test would demonstrate that the test scores obtained on computerized tests were not significantly different than scores obtained on paper and pencil tests.

3) The third hypothesis was that CAT would allow for significant increases in test efficiency as measured by the time taken to write a given test, when compared to paper-pencil tests.

Limitations of the Study

Due to the nature of the teaching environment where the study took place, there are a number of limitations to the study that should be noted. First, as the software design and development of the CAT were undertaken in the Macintosh environment, design approaches used in this study may not be directly applicable to DOS environments.

Second, because the students involved in the study were adults, the findings may not be generalizable to other age levels. Specifically, the results may not apply to adolescent and pre-adolescent learners due to differences in motivation and experience. In addition, differences in experience with this type of item format would also contribute to a limitation of the study.

Finally, because of the self-paced nature of the course, tests were not administered simultaneously to all students. This created the possibility that some students might receive information about a test from students who had completed the test. However, this possibility was common for both the computer and paper administered versions of each test.

Importance of the Study

Several trends have contributed to the increasing focus on computers within the educational system, including a shift towards individualized learning, an appreciation for the skills and attitudes of students, and an increase in computer use by society as a whole. The increased focus on student centered learning has stemmed in part from the recognition that adults learn at different rates and use different learning styles (Knox, 1986). This realization has lead to an increase in what is termed resource based learning and has been evidenced by the increasing use of resource centers and of individualized learning packages.

The goal of student centered learning is to provide the student with a flexible system of learning geared to their individual learning needs. The role of the teacher is to provide a supportive rather than central role in this learning. Student centered teaching uses tutors, study guides, laboratories, and a variety of other strategies to meet the needs of the individual (Percival & Ellington, 1984). This trend towards student centered learning is also reflected by the increased appearance of distance education and open learning centers. More adults are recognizing that they must continue their studies to gain the skills needed to survive in an increasingly technological society. Open learning centers offer programs within a self-paced format and use a wide variety of different instructional and information sources. Students may take a range of offerings from basic college upgrading to university transfer courses.

There are a number of advantages to using an individualized learning approach. First, the coursework is usually more accessible to students. As learning is self-paced students are not rushed past difficult concepts nor are they held back on the easier concepts. By including learning objectives with each unit of study, students can directly focus their learning towards a subsequent test or examination. Many student learning modules have also incorporated a variety of different media sources. In this way the characteristics of each medium are used to create a learning environment that best suits each student.

The disadvantages related to the individualized approach relate to inadequate student commitment and motivation. Younger students may lack the strong goal orientation necessary to make the system work effectively. In addition, teachers must also adapt to a new, more facilitative role within the educational system and many teachers may have difficulty in adapting to this type of approach. Finally, while the individualized approach relates well to courses with a strong factual content (i.e., science and math), it does present problems for student assessment.

Another factor influencing the extent to which computers will be used is teacher attitudes towards the technology. If educators are to introduce substantial changes by computer technology they must be comfortable with that technology. Teachers who view computers as a threat to their jobs, authority, or traditional role may fail to provide the cooperation for such innovations to succeed. This seems to underscore the need for effective professional development to complement the investment in hardware and software.

A final trend would seem to be the recent developments in computer hardware. With the arrival of inexpensive microcomputers it is reasonable to assume that most institutions will have increased access to them. As computers become increasingly more sophisticated they have become smaller and more compact with more memory and faster processors.

The types of educational software now being produced are of much better quality than in the past. Much of this is due to the recent developments in hardware and to an increase in the number of software titles available. Educational software is becoming comprehensive, flexible and easy to use. The arrival of many authoring packages (i.e, Hypercard) is also increasing the amount of software available.

The goals of the Year 2000 initiatives proposed for the public school system points to a future educational system characterized by organizational, curricular and assessment flexibility (Province of British Columbia, 1990). The goals of the Graduation Program include encouraging students to learn independently. Students will also be expected to select and apply appropriate learning strategies and to learn at their own development rates. Students will be exposed to a number of self assessment methods and evaluative techniques and will also be involved in determining the method and timing of their own assessments. All of these clearly point to a more student centered approach to learning in the future.

These initiatives also propose to increase the accountability of the system through more detailed reporting. This may effect great changes in the way that courses are offered. Teachers will be required to provide both instruction and testing on demand for students. Some college settings currently offer courses in a similar individualized format.

The current direction of educational change has precipitated a clear increase in the educational use of computers, and the range of educational uses for computers has also expanded rapidly over the last decade. This increasing use of computers requires a broader look at the ways that computers can be used as effective tools in education and training.

Computers were first used as calculators (Percival & Ellington, 1984). This use developed from the need for academics to do complex and repetitious calculations as part of their research. Another use involved the manipulation of formulas to model situations in science and engineering. These uses set the stage for more diverse uses for the computer in education. As computers became more widely used for research, the need for education in the use of computers and in computer programming increased. It was the inclusion of computer programming in the high school and college curricula that introduced these institutions to computers.

Computers have been used as a direct aid to the teaching process. The use of computer-assisted learning (CAL) has made a major contribution to education and training. In CAL the computer acts as a tutor with which the student can participate in an interactive dialogue. Examples include textual, graphic, and other types of information. The great flexibility and data-handling capabilities of computers allows CAL to adjust to the needs of the individual student.

More recently, computers have also taken on an administrative or managerial function in education. This area is known as computer managed learning (CML). In CML the computer acts as an administrative support to the teacher. Some variations of CML provide individual guidance and assessment to students while at the same time providing clerical support by maintaining records and generating reports.

With educational personnel increasingly more reliant on tests and other assessment devices to provide data for student diagnosis, placement, selection, feedback, and evaluation, the use of computers for this purpose has steadily increased (Mizokawa 1991). This new technology has developed into the field of Computer Managed Testing (CMT). CMT usually features multiple-choice (MC) type test items. The advantage is that items can be stored in large banks and can then be selected randomly or by specific characteristics. Publishers have included such item banks with textbook orders. Teachers are allowed the flexibility to customize their tests and present a more professional looking product.

The use of the MC testing format is a frequently used technique for evaluating student educational or psychological performance (Burns, 1990). The high reliance on MC testing alone does not support their continued use. There are a number of distinct advantages and disadvantages to the use of MC tests and it is important to address some of these here.

One advantage of MC tests is that they are marked with complete inter-marker reliability (a score on a given test would be identical given another scorer due to the objective nature of these items). Also, large numbers of questions can be answered ensuring a thorough sampling of the course objectives. Test items can be designed to test specific outcomes and these items can be trial tested to determine their difficulty level. MC tests lend themselves to easy marking and to thorough statistical analysis. This allows for investigation of individual difficulties as well as general problems in the student population. MC items are also difficult and expensive to construct and therefore require considerable preparation time. Their apparent ease of construction has lead to many amateurish attempts at constructing MC items, the result being many poor and invalid items. Another disadvantage to this type of item is that the teacher or marker cannot see the reasoning behind the choice of an incorrect answer. Finally, it is often difficult to write MC items that assess higher level abilities.

CMT applications, hold some distinct advantages for educators. The programs allow their test items to be randomized thereby allowing for flexibility and variety in the testing situation (for example different but equivalent tests can be designed by a computer by the random selection of test items selected by specific learning outcomes). A teacher may also print multiple versions of identical tests by scrambling the order of presentation of items. In addition, most programs allow for on-going analysis and editing of the test items. Editing features allow the teacher to remove poorly worded

items or to add new items to a test bank. Item analysis features can allow a teacher to tailor the difficulty level of a given test.

As the use of computer managed testing programs has become more popular this area of assessment has developed to include an area known as Computer Administered Testing (CAT). This type of application has maintained many of the features present in CMT but in addition serves to directly deliver these tests to the student, therefore eliminating the need for printed forms and answer sheets. A CAT could consist of a series of MC items presented in linear fashion by the computer.

Computerized testing programs have been demonstrated in a wide variety of areas including adaptive testing (Burns, 1990) and psychological testing (Schuldberg, 1988). In addition, many colleges and university science programs have CAT programs in place for evaluative purposes (Fletcher & Collins, 1985). The administrative benefits of CAT are many. Computerized testing allows increased flexibility in testing and supports a more individualized and student centered approach to teaching. However, before the use of CAT becomes widespread it is important to ensure that students are not disadvantaged by this new format. Careful consideration must be given to the related issues of software and interface designs, test format equivalence, and student attitudes towards this format of testing. These areas provided the focus for this study.

Chapter Two

Literature Review

The questions addressed by this study centre on two areas in the current literature. The first area sets the context for the study within the areas of adult education and assessment. The second relates issues of software and interface design to the development of computer administered testing and addresses the comparative merits of computer administered and written tests, including student attitudes related to each of the testing formats. These areas are interrelated, but for convenience they are considered separately here. Before the issue of computerized assessment can be addressed it must first be placed within the context of adult education and assessment in general.

Often educators are not convinced of the differences between teaching adults and children (those of andragogy and pedagogy). The growing knowledge base of the characteristics of the adult learner is highlighting these differences (Barer-Stein & Draper, 1989).

Voluntary learning has long been used as a distinguishing feature of adult education. The voluntary attendance of adults at classes highlights a fundamental departure form the normal school-based education. Because the adult voluntarily enters into the educational transaction with the teacher, the nature of the educational process is significantly different (Brookfield, 1983; 1986).

Knowles (1970) and Darkenwald and Meriam (1982) discussed what they believed were the four assumptions of andragogy. These began to define the adult learner and are listed here:

1) that as a person matures their self-concept moves from one of a dependent personality toward one of a self-directing human being;

2) that an adult accumulates a growing reservoir of experience, a rich reserve for learning, (for an adult, personal experiences establish self-identity and so are highly valued);

3) that the readiness of an adult to learn is closely related to the developmental tasks of their social role; and

4) that there is a change in time perspective as individuals mature, from one of future application of knowledge to immediacy of application; thus an adult is more problem-centred than subject-centred in learning.

Considering these developmental traits adult education assumes that students are functioning members of society (a non-preparatory approach). As adults make daily decisions affecting their lives it follows that they are capable in the planning and implementation of their own learning. The shift in emphasis in adult education (from preparatory learning to the enhancement of adult living) warrants approaching adult learners as individuals (Knowles, 1970; Darkenwald & Meriam, 1982). Darkenwald

and Merriam go on to iterate that the effectiveness of adult educational programs often depends on a thorough understanding of the individual needs, problems and attitudes of adult students.

Knox, (1983) has stated that adult learning can also be affected by a variety of different circumstances specific to the adult learner. These are tied to the development of the adult. Characteristics that may influence the adult learner include the following:

1) The physical condition of the learner. Physiological condition, sensory impairment, and physical health can effect learning and cognition in various predictable ways;

2) The adjustment of the learner. The effective accessment or facilitation of learning is less likely when their is substantial personal or social maladjustment in the learning situation. This type of situation is characterized by learner defensiveness and anxiety;

3) The relevance of the curriculum. The adults' motivation and cooperation in learning is more likely when the tasks are meaningful and of interest to the learner;

4) The speed and timing of the learning. Time limits and pressures tend to reduce learning performance in adults. Optimal performance is more likely when learners can proceed at their own pace;

5) The social status of the learner. Socioeconomic circumstances are associated with values, demands, constraints, and resources that can affect learning ability;

6) The changing circumstances of the learner. Social change can create substantial differences between older and younger adults regarding their experiences and values (this is often described as the generation gap);

7) The personal outlook of the learner. Personal outlook and personality characteristics (such as openmindedness or defensiveness) can affect the way that an adult deals with specific learning situations;

Adults engage in learning activities mainly to enhance their performance (Knowles, 1970). This can include a combination of knowledge attitudes and skills. An understanding of the discrepancy between current and desired proficiency is the goal of most assessment practice in adult education (Knowles, 1970; Knox, 1981; 1986). Awareness of this discrepancy can motivate an adult to engage in more active learning.

Adult education also needs to respond to learners' diversity. As people grow older they become more and more different due to the accumulation of individual experiences (Knox, 1986). Many adults prefer active roles as users of education rather than recipients of education. This preference for active involvement reflects the self image of the adult learner as that of an independent person. One way to be responsive to adults'

preferences for active participation is to include them in the objective setting process. Knox goes on to explain that learner involvement in evaluation and assessment is also important. Self assessment and self-paced learning helps learners to become more independent, creative, and self-reliant in future self-directed learning activities.

The lack of an individualized program for adult instruction can also form a barrier to participation in adult education. Various institutional barriers account for between ten and twenty five percent of potential adult learners according to most surveys (Rubenson, 1983). Institutional barriers include the following five general categories: scheduling problems; problems with location and transportation; lack of courses that are interesting, practical or relevant; procedural problems related to attendance, time requirements, etc.; and a lack of information about procedures and programmes. The most commonly reported barriers are scheduling and problems with location or transportation.

The theme of individualized learning can be continued into the area of assessment. Further research in adult education has shown that student referenced evaluations or assessments are the most conducive to optimum learning. Much of this stems from the idea that owning one's strengths and weaknesses is a prerequisite for learners moving out of a dependent stance (Barer-Stein & Draper, 1989). Here, the degree of independence is not signalled by the physical context in which the student is working but by the

degree of control exerted over the content and method of learning or assessment (Brookfield, 1983).

Before the suitability of an assessment can be determined the purpose of that assessment must first be considered. While most educators use assessment or evaluation purely to determine classroom achievement or grades in a summative manner still other types of assessment are of a more formative nature, including self-evaluation for remediation or review and college placement and proficiency exams. The desirable characteristics of any assessment method would include the issues of test validity and test reliability. In addition, assessment methods should be practical for both teachers and students (Percival & Ellington, 1984).

One use for an assessment tool is to guage the amount of significant learning that takes place outside of the formal education setting. There is an enormous amount of this type of learning taking place and it occurs in both individual and collective settings (Brookfield, 1986). Brookfield also noted that since adults bring a collection of individual experiences, skills, and knowledge to the learning situation. This should warrant a more individualized approach to their assessment.

The type of assessment best suited to the type of student found in adult basic education at the community college level is closer to the formative model (due to the nature of the adult student). While achievement testing is done, students are first evaluated in an effort to place them at an appropriate

level to begin their upgrading and then repeatedly re-assessed in order to determine appropriate areas for remediation or review. As is common in the teaching of adult education, these assessment strategies must be flexible and student centred (Moore & Waldron, 1981).

Assessment tools may be required on short notice and must be readily available to students at the college. The adult student has many time commitments including children, jobs, etc. and are balancing these responsibilities with their role as a student. If assessment is not convenient for them, it may represent a missed opportunity. The environment and design of the assessment tools must be clear, non-threatening and practical. Due to the user-pay nature of most college up-grading courses, students will not remain in a situation in which they are uncomfortable (Moore & Waldron, 1981).

In general, the design and use of computer administered tests must be considered both in the context of adult students and in the formative nature of the testing environment. In this, the design of a computerized test may effectively place limits on an assessment strategy or conversely precipitate greater flexibility and reliability in the testing environment.

Mizokawa (1991) described the process of both designing and implementing a computer administered testing program. The project was conducted at the University of Washington and was intended for the public school system. In developing a computer administered test design for

language development, the project team lay out a series of recommendations for the implementation of computerized testing programs:

1) that during the selection of the hardware requirements, the team assume that the project would run on a bare bones system and that the available hardware was in the project schools before programming was done;

2) that the design team consider human design factors in the administration of the tests. This included the application of psychometric principles at all times during the construction of the tests (inferring that the tests were designed using the same concerns as in conventional test formats);

3) that the programmer minimized long pauses for the recording of accumulated data, instead recording data more frequently with shorter durations of disk activity (this guideline eliminates "dead time" where students could become bored and randomly push keys);

4) that the project team stated that task designs be as simple as possible so that the test taking behaviors would not interfere with an accurate demonstration of a students knowledge or ability;

5) that the test instructions provided should be legible and simple in content and readability (i.e., using large and legible fonts) and that in addition these directions should be given at a grade level less than that of the readers;

6) that the software should provide demonstrations of the expected behaviors (i.e., the use of animation to engage the users attention) and that students would be allowed to skip or omit these instructions if they wished;

7) that the computerized tests should be administered in more frequent but shorter sessions (the rationale included issues related to comfort and fatigue contributing to errors) an that the need for students to be adequately trained in test-taking behaviors;

In summary, most of the guidelines developed by the team at the University of Washington could apply to computer managed testing programs of any kind. Mizokawa pointed out that computer administered tests should be designed with similar consideration given to units of computer assisted instruction (CAI). These include adequate preparation, pre-training and on-line assistance. These and other related issues were also discussed in an earlier published guideline by the same authors (Mizokawa & Hamlin, 1984).

Burns (1990) designed and implemented a computer administered testing program at the State University of New York. The project was designed to provide a multiple-choice diagnostic test design. It was intended for use with learning disabled and physically handicapped students. On completion of the project they discussed several general factors that can be considered when designing diagnostic assessments for use with microcomputers.

The first factor that they considered was that of identifying the content area to be evaluated and ensuring its suitability for the computer format (i.e., no long text passages, adequate screen display, etc.). The amount of available random access memory (RAM), may also limit a specific area of content in a test due to the use of graphics or sound.

Referring to test item characteristics, Burns related that multiple-choice items are composed of a stem and a series of alternatives. The student task is in selecting the best of the available alternatives. Burns advised that if a "none of the above" alternative is used it should be viable to the extent that it is sometimes the correct answer. Finally, in some situations a presentation of a series of sample items will be necessary to ensure that the student understands the task and the method for selecting alternatives. These sample items should be low in difficulty.

Burns continued by outlining another popular design feature that allows test items to be selected randomly for presentation from a pool of available items. The items may be selected using some form of random sampling or sequentially. This allows for many different test configurations, thereby limiting the possibility of fixed order effects on repeated tests.

Finally, Burns related that for computer administered tests feedback is often an important component of multiple-choice computer assessment. The results component is critical to every multiple choice computer assessment.

These may consist of summary and item results. Summary results may include the number of items, the number correct, percentage correct, and time required. The item results may include a listing of the items presented, the correct answer, the incorrect response, and the response time. All results data can be stored on disk for inspection later.

Summarizing, Burns concluded that although multiple choice tests are often criticized (especially the normative interpretation of scores) the design and development of multiple choice computer assessments can result in useful information for better understanding student performance.

The recent introduction of the software development tool Hypercard by Apple Corporation and the inclusion of this software with the purchase of Macintosh computers has lead to increases in the production of educational software. The ease of programming in Hypercard and the simplicity of the user interface has some important implications for both computer assisted instruction and educational research. The Macintosh is among those computers most widely available on university campuses (Bowers & Tsai, 1990).

Bowers and Tsai (1990) give a brief overview of the Hypercard program and its potential implications for future educational research. Hypercard is very briefly described as a software development environment that contains: a development language called Hypertalk, a set of tools for creating and editing graphics and text, and the means for creating and saving

stand alone programs called stacks. This type of application is generally known as course authoring software.

The type of programming which Hypercard employs is often termed "object oriented programming." This environment operates using a series of "system messages," (e.g., the message "mouseDown" is sent when the mouse button is depressed). System messages are either intercepted by a variety of different "objects" (i.e., as stacks, cards, backgrounds, fields and buttons). Reception of a system message causes the execution of program sub-routines called "scripts" that are contained within the given variety of objects. Students interface with the program by clicking the mouse on these objects. System commands within Hypercard operate along an "object hierarchy," (i.e., from the button or field level through the card, background and stack levels). Bowers and Tsai believed that such a system could facilitate tracking the types of student behaviors elicited during the use of a particular stack (i.e., time spent on a card, etc.). They believed that the flexibility of Hypercard and it's Hypertalk language has implications for the educational researcher.

The authors described a case study where Hypercard was used as a data collecting tool in Bowers' dissertation research. Here they emphasized the technical aspects of using the Hypercard program rather than the actual research involved. The research implementation of Hypercard was largely used in the collection of learning path data by innovative scripts (i.e.,

program sub-routines). The researchers' program generated summary reports for each subject that were later analyzed.

The authors relate that Hypercard can be successfully used to collect, report on, and analyze learning path data. They summarized by stating that this type of research application in combination with qualitative methods may yield insights into cognitive and learning processes.

Sarti (1992) reviewed the architecture of educational applications and of the human-machine interface. Sarti noted a general trend towards easy to use, ergonomically designed interfaces such as those using the object oriented paradigm. The author discussed the evolution of this interface through various stages including the teletype interface - a text only interface (characterized by screen dialogues such as "press return to continue") to the more recent "window" interfaces allowing the computer several active areas on one screen.

Sarti identified a number of advantages to the use of window managed computer systems. First, he found this interface easier for users to understand as it was related to in a direct way (by pointing with the mouse). He also found that the use of icons allowed users to work without learning a formal command language for the operating environment. Sarti stated that the combination of these aspects contributed to the shorter time required for training individuals in the use of these types of programs. Finally, Sarti discussed the object oriented environment as a useful tool in the design of

educational software. Sarti observed that the manipulation of programmed objects was a natural way to design a user interface.

The comparative merits of computer administered versus written tests have also been discussed at some length in the current literature. Areas of interest include potential time savings during test administrations, varied and increased student feedback and possible cost savings for educational administrators. Other important questions relate to the type of computer based test and the amount of flexibility in the test-taking behaviours allowed by each system design.

Computer based tests have been developed to assess placement level, evaluate achievement, provide for remediation and review, and to administer psychological tests. Common issues related to the use of computer-based tests in all areas include test reliability, validity, and equivalence of results with typical paper-pencil type tests.

The potential benefits for computerized testing programs have been discussed in reference to adult training programs (Feuer, 1986). Feuer pointed out that while assessment was initially an individualized process, this has changed due to the need for widespread testing. While large scale testing programs exist these can pose many problems for evaluators in terms of test security, test scheduling, and delays in reporting results. The computer administered format of testing was discussed by Feuer as a potential solution to these and other problems.

Feuer outlined the organization of PLATO, a computer based testing and training program. The immediate feedback feature of such testing programs was viewed as allowing a return to individualized testing where assessment was available on demand. The administrative features for record keeping and report generation were seen as a great benefit for the test administrator owing to the computers extensive data retrieval and storage capabilities. Feuer continued by discussing other potential forms of computer assessment that could include computer simulation.

Feuer also discussed computerized adaptive testing as another potential benefit to the computerized format of testing. Computerized adaptive tests are seen as intelligent tests as the computer scores the student responses as the student progresses and decides which items to present next based on an estimate of the student's ability. The intent of such tests is to zero in on the precise skill or ability of the examinee. Feuer believed research with computerized tests has shown them to take less time than conventional tests and to provide more precise proficiency estimates.

Olsen (1990) studied three different modes of testing; paper-administered, computer-administered, and computer-adaptive tests. Computer adaptive tests differ from other computerized tests in that the difficulty level of the test items is adjusted according to a student's performance on previous items on the test. This allows the computer to zero in on a student's ability level. The participants in the study consisted of 350

grade 3 students and 225 grade 6 students from three California elementary schools. The test items used were mathematics applications items selected at random from the California Assessment Program.

The results of this study demonstrated high levels of compatibility among test scores obtained with the three different testing formats. In addition, significant time savings were noted for both types of computerized assessment. In the computer administered form of the test, the grade three students completed the computerized tests on average 42 per cent faster than the written format, while the grade six students completed the computerized test on average 22 per cent faster.

In another related study (Olsen, 1990), the use of computerized testing was related to student aptitude testing. This study involved seventy-two grade six and seven students from Utah schools. The sample consisted of an equal number of boys and girls taking a battery of fifteen tests related to the following aptitude areas: verbal, quantitative, visual/spatial, reasoning, memory, and attention.

The results of the study indicated significant validity correlations between the computerized and written formats of the tests. The results of the two studies support the idea that computer administered forms of testing require less time to administer than conventional tests while providing comparable results.

Many psychological tests have been converted to computerized formats and studies have been directed at showing the equivalence of these computer administered tests. Researchers at Texas A & M University studied one such version of the Strong-Campbell Interest Inventory (Vansickle, 1989). The participants in the study were seventy-five first, second or third year university students enrolled in an introductory educational psychology course. Students were selected at random from four sections of the above course and placed in one of four treatment groups. Administration of the tests was conducted both at the beginning and end of the course with computer-based tests given by appointment and the written format given during group sessions.

The main purpose of the study was to examine test-retest reliability of the computer testing format. Procedurally, this was accomplished as follows: the first group was tested and retested using the written format; the second group was initially tested with written tests, then retested by the computer; the third group was tested by the computer, then retested with written tests; and finally the fourth group was tested and retested using only the computer.

Using statistical significance the results of the study indicated the two forms of test administration were equivalent. The researchers pointed to similar test-retest reliability co-efficients, means and standard deviations to support their arguments. Interestingly, the group tested and retested with the

computer format of the test exhibited the highest test-retest reliability of all the treatment groups.

In another study at Texas A&M University (Kapes & Vansickle, 1992) a comparison was made between computer-based and paper-pencil versions of the Harrington-O'Shea Career Decision-Making System. The study focused on the equivalence of mode of administration for this specific interest inventory, pointing to similar studies which have yielded conflicting results. The researchers asserted their belief that the use of computers in assessment does not produce results that are significantly different from conventional test administrations.

The participants in the study were sixty-one undergraduate students enrolled in a section of an introductory secondary education course at Texas A&M University. The study was conducted as a two-week test-retest repeated measures design. Students were randomly assigned to one of two equally sized groups with one group tested and re-tested using only the computerized format and the other using only the paper-pencil format. The two groups were then compared across a number of different measures. A comparison of the mean and standard deviations between the two groups yielded no significant differences but the computer-based format of the test yielded significantly more reliable results than the paper-pencil version on three of the six scales measured. Additionally, the computer-based group completed the tests in a shorter time period. The researchers argued that these results demonstrated the equivalence of the computer-based version of

this test, pointing to the increased reliability of the test and potential time savings in its administration as further advantages.

A further utility for CAT is for the purposes of remediation and review. A study at Memorial University demonstrated that CAT was useful in identifying areas for remedial teaching (Collins & Fletcher, 1985). The study investigated the use of computerized tests as a tool for diagnosing areas for remediation and review in an introductory university biology course. The experiment was conducted with two classes of a non-credit, one semester course and was structured so that a multiple-choice computer test was taken by each student individually after completion of each course unit. Course units were studied individually using behavioral objectives and a textbook. Tests were taken on an Apple computer with the students given detailed feedback after each test.

Following this, instructors analyzed the incorrect responses on the tests to focus their remediation classes. Corresponding written tests were then administered on the same material at the end of each remediation session. The written test scores were then analyzed by comparing scores on revised items to scores on the non-revised items taken from the same test. The results of the study indicated that for the revised items, the gains in scores between the computer and subsequent written test were greater (i.e., half of the students who previously had the item incorrect were now getting it correct). They concluded by relating that the use of computer tests was significant in facilitating diagnosis and remediation in Science Education.

In another more recent study at Memorial University (Fletcher & Collins, 1987) researchers sought to determine whether first year university students were disadvantaged by computer administered testing in Biology. In this program, computer tests were used in place of traditional paper-pencil type tests. The tests consisted of multiple-choice items.

The study consisted of two components: the first compared student achievement scores on identical tests delivered in the computerized and written formats; the second analyzed student responses to a written questionnaire. All students in the study first completed two tests in the computerized format to familiarize them with this type of test. The students were divided into two groups with each group alternating between the written and computer test formats. The researchers recorded test results for both groups over a series of eight tests. At the end of the study each group had completed four computer and four written tests.

Finally, the scores on the two formats were compared. The results showed no significant difference between the achievement scores on the two test formats. The student responses to the written questionnaire allowed the researchers to categorize the relative advantages and disadvantages for both written and computer administered test formats.

Referring to the computer tests, students listed the following as the greatest advantages: the immediacy of scoring, the ability to receive

immediate feedback, and the convenience and speed of computer tests. The students also listed the following as major disadvantages of the given computer format: the inability to review or change answers and the inability to skip questions. Concerning the written tests, students listed the following as the greatest advantages: the ability to change answers, the ability to skip questions, and the possibility of obtaining hints from other questions. The students listed the following as disadvantages to the paper and pencil method: the delay in obtaining feedback, the possibility of changing correct answers, and the slower delivery of the written tests.

Tait (1984), in a study at the University of Leeds, related another experience with a computer-based learning system. The study described the merits of a computerized MC testing system that was used as a self-teaching and self-assessment tool for first year pharmacology students. The focus of the system was as a learning diagnostic aid for students with student feedback provided as an important part of the program. On completion of the study the researchers found a high correlation between the hours spent by students using the computer testing system and their achievement on the MC test portion of their final exam. On average, students spent nine hours using the system and were reported enjoying its use. The researchers suggested that the system could provide additional remediation and direction for students during exam preparation.

Despite the support for CAT, a number of studies have yielded ambiguous or inconclusive results around the potential benefits of this

method of assessment. In an attempt to gauge individual sensitivities to computerized tests Schuldberg, (1988) created a sensitivity score that measured an individual's sensitivities to computerized test administration and repeated testing. For this, Schuldberg used a common psychological test: the Minnesota Multiphasic Personality Inventory (MMPI).

In this study measures of individual sensitivities were obtained by counting the number of changed items on subsequent administrations of the computerized test. Schuldberg discovered that the measured sensitivity scores tended to be largely time sensitive rather than format sensitive. However, individual items did tend to change with repeated testing in the computer format. Although no suitable explanation for the phenomenon was offered, Schuldberg pointed to software design issues rather than to the overall format of the test.

Waugh (1985), used computer diagnostic testing coupled with mastery learning programs. In a study at Georgia Southern College, the effects of CAT on science achievement and attitudes was documented. Researchers followed a diagnostic testing strategy which involved the development of behavioural objectives and objective-referenced test items

The study group consisted of 36 ninth and tenth grade students enrolled in three biology classes at a Georgia high school. The students were randomly assigned to either a treatment or comparison group. During each of the treatment sessions students were given behavioural objectives

and required to schedule a 30 minute outside-of-class diagnostic testing session on the computer. The comparison group was given another assignment requiring an equivalent time commitment.

At six intervals during the study student achievement and attitude were assessed. The results of the study were conclusive, revealing significant differences between the treatment and comparison groups in achievement and attitude (favouring the written format). Waugh attributed these results to the relatively small sample in the study rather than to the overall effectiveness of the different testing formats.

The effect of the computer medium on examinee performance and test item characteristic was also considered in a study by the American College Testing Program (Spray, Ackerman, Reckase, & Carlson, 1989). In this study design features of several computer administered tests were compared. The authors point out that past studies had yielded unequivocal results among computer-based and written versions of various achievement tests (Moreno, 1985).

Spray, et al. hypothesized that if computer-based versions of a test had been sufficiently flexible, allowing examinees to review or change answers, these disparities would not appear. Spray, et al., continued by outlining a computer-based testing system on IBM computers in which the range of possible test-taking behaviours was as closely matched as possible with the

written format. Their study involved students enrolled in the Marine Corps Communication-Electronics School, at Twenty Nine Palms, California.

The basic design of the study was to involve examinees randomly assigned to a treatment group (113 students) and a comparison group (121 students). Each group was administered a 25 item objective test on three occasions. Test items were randomly selected by item-objective from a pool of 75 test items. The results of the study indicated that there were no significant differences between the different test media during the first two sessions. However, on the final administration, there were significant differences found favouring the pencil-paper format. On this test students scored on average higher than those taking the computer-based test. Spray, et al., pointed to specific items included in this final test to account for the observed differences in achievement scores.

This brief outline of the current literature underscores a continuing need for research in the area of software design as it applies to computer administered testing. Further, more conclusive evidence is required on the reliability and validity of computer programs as testing devices.

Chapter Three

Research Overview

The goals of this research project centred on three different aspects of CAT. The first aspect related software and interface design issues discussed in the literature to the design and development of a prototype CAT application. The second aspect of the study related the comparative effectiveness of this CAT application through the comparative analysis of test scores obtained by the computer and traditional written tests. The study attempted to subjectively gauge student attitudes towards the computer administered and paper-pencil formats of testing with pre and post surveys. Each aspect is discussed here in greater detail.

Background of the Study

The context of the study was a learning centre at a small community college. The centre administers grade 11 and 12 secondary science courses within the context of Adult Basic Education. The curriculum taught included Biology 11 and 12, Chemistry 11 and 12, Computer Studies 11 and Physics 11. Instruction was individualized with all courses offered in a self paced learning format. Students were encouraged to work through their required course units and were assisted with behavioral objectives and individual personalized instruction. The evaluation for each course was a series of MC tests complemented by a variety of lab and written assignments. Student assessment was provided on demand.

Due to the nature of the program, a number of attempts were made to enhance and streamline student tracking and evaluation. This led to the consideration of several computer administered testing packages. One of the noted concerns was in the inflexibility of many of the available testing applications. Most of the computer-based tests evaluated were difficult to learn or of inflexible design. For example, applications often did not allow students to skip questions or to review or change their answers on a question. The potential impact of this type of inflexibility on student test performance has been related to poor performance on achievement tests (Moreno, 1985). It was in response to these concerns that the possibility of developing an improved computer administered testing design was first considered.

Over a period of four months students were informally interviewed regarding their performances on the multiple choice tests (tests were often marked with the student present). During these sessions, many students reported difficulties in answering typical multiple choice test items. This was despite students receiving instruction in proper test taking strategies. A frequently reported claim of students was that they were easily distracted by the alternative choices (or distracters) posed by a question. On consideration of these and other student concerns, a decision was made to address these issues through the design of a more student-focused computer testing application.

Design of the Program

The advantages seen in designing a computerized testing program included the following. First, test items could be presented one at a time in a linear fashion allowing the student to consider each item individually without distractions from other items. The possibility for immediate scoring would allow for more efficient student tracking and evaluation while allowing immediate feedback to students, (tests would be scored on completion of a test). The use of advanced feedback features could be included for assisting students in remediation and review. The possibilities for this type of feedback included providing students the ability to review their incorrect responses and to allow them access to other summative information such as the time taken in completing a test and their final score.

Most of the documented disadvantages to CAT have been related to issues in software or interface design (Moreno, 1985, & Spray, et al., 1989). In considering this, the resulting computer testing application was designed with the aid of considerable student input and additional information from the recent literature in this area. A decision was made to pilot test the proposed application before beginning the major portion of this study. The intent was to gain additional student input regarding the test design and to further analyze the system requirements.

The development environment chosen for the purposes of this study was the object-oriented Hypercard authoring language running on a Macintosh computer. This "windows" type environment is easy to use and requires minimal training in the system operation. User actions consist of pointing or clicking a mouse at various screen icons. The flexibility and ease of use of this environment were instrumental in its selection.

Throughout the design process relevant software development guidelines and test design principles were followed (Mizokawa, 1991; Burns, 1990). The general test layout included introductory and summary directions for students as is common practice in the design of written tests. Considerable care was used in keeping the design of the student interface simple or "user-friendly." The resulting program has incorporated many features similar to paper and pencil type tests while retaining the noted advantages of the computerized format.

At the start of the program, the computer asked for the student's name and would use this to further personalize student feedback. The basic layout of the test design was that of an "electronic page" whereby items were presented one at a time. An attempt was made to increase the legibility of the included text through the use of a larger font and the careful layout of the various text fields. The number of items on each test was limited to twenty in an attempt to avoid student fatigue during the testing sessions.

The following is a list of design features incorporated in the initial version of the testing prototype (two other features were added following the pilot study). A number of screen-prints and scripts are included in appendices A and B to further illustrate these features.

First, students were provided with on-screen "buttons" that allowed them to move between pages or items on a test and to access a tutorial. The nature of each of these buttons was further indicated by the choice of icon used. Arrow icons were chosen to enhance the intuitive nature of the navigational buttons. The locations and type of the various buttons were fixed between screens and between tests to provide both consistency and uniformity in the testing environment. The number of buttons was also limited to simplify the interface.

The use of navigational buttons also allowed students to review or change their answers as they moved backwards or forwards through a test. Students could elect to skip over questions and return to them later on in the test (this feature is not often included in computer-based tests but is often listed as an advantage of the paper-pencil format of tests). The provision of this feature was provided through the development of a script which scores and rescores an item each time it is accessed by a student. A student could indicate the correct answer by clicking in its adjacent box. Answers were automatically changed by clicking beside another option.

Another feature incorporated into the test design was the provision for immediate scoring. On the last item of each test a button was provided which could autoscore the test. This feature was allowed through the development of a script that scans an entire completed test summing the items marked correct. Students are then presented with a pop-up field displaying their name and final score.

Informal discussions with students also led to the inclusion of another design feature in the program. This feature was termed the "elimination feature" as it allowed students to eliminate multiple-choice alternatives from their field of view. The rationale here was to allow for greater implementation of a "process of elimination" test-taking strategy on the computerized version of a test. This idea is similar in nature to students stroking out incorrect alternatives while completing a paper-pencil test.

On the computer-based test students could use the elimination feature to eliminate incorrect alternatives by simply pointing at these alternatives with the mouse and clicking a button. The selected option would then be hidden from the student's view. This process of eliminating alternatives was easily reversed by clicking the mouse on vacant locations on the screen left by the hidden fields. Eliminated options were automatically displayed to the student when the item was re-presented (i.e., during a review of previously answered questions).

Description of the Pilot Study

In March (1992) a small pilot study was conducted. Specifically, the study was completed using a computer lab running from a networked fileserver. There were two purposes to doing this preliminary study: the first was to obtain additional student feedback regarding the design and interface of the computer testing program; the second was to provide some preliminary data collection and analysis. However, information gained through the course of the study contributed to further revisions in both the software design and in the organization of the research method.

Seven volunteers were selected from a class of 14 adult students enrolled in an introductory computer studies course, (Computer Studies 11) within the context of Adult Basic Education. None of the students had previous experience with Macintosh or other computers before this introductory course. Achievement scores for these students were recorded for an earlier paper and pencil multiple-choice test that had been administered to the entire class. Volunteers for the study came forward on the basis that their score on a computer administered test would replace their earlier test score if they were to score higher.

The students then completed a computerized multiple-choice test that assessed the same educational objectives and consisted of the same number of items as the earlier test. The testing was conducted in the computer lab during one group session. On completion of the test it was computer marked

and students were given immediate feedback regarding their scores. The students were asked a number of questions during an informal interview session. The analysis of student response patterns and test scores yielded some useful information.

Students were also asked to respond to questions relating to the format and general layout of the test. When asked whether they found the format distracting, six students indicated that the interface design was simple and did not cause distractions. When asked about the elimination feature, four students reported that they had used it while three students reported that they had not. When asked about the flexibility of the test, all students reported using the buttons to review or change their answers. All students reported a preference for the instant scoring feature on the test.

Curiously, the mean achievement score of the test group on the earlier paper version of the test was identical to the overall class mean on that test of 65.7 percent. However, test scores on the computer version of the test yielded a mean value of 61.7 percent (four percent lower than in the paper-pencil version of the test). Further analysis of the test scores using a t-test showed that the variation between the two groups was not significant.

The small variation in the test scores can be explained. When students were asked whether they had studied for this test, five students reported that they had not prepared in any way, while two students reported that they had adequate preparation. Certainly, the design of the study did not optimally

motivate students to perform. As the test items were not identical in each version of the test, the computer-based test may have been at a higher difficulty level.

Students preferred the computer version of the test to the earlier paper-pencil method but they did suggest the inclusion of an on-line help system that would demonstrate the features available during a given test. Students also indicated that a review session provided at the end of each test would be helpful. Both recommendations were eventually incorporated into the revised test design before the main study.

Following the results of the pilot study, an effort was made to refine certain elements of the software design. While the basic layout and interface remained unchanged, two additional features discussed with students during the study were also included in the revised test design. Continuing feedback from students also led to minor adjustments in the design of the user interface.

Revisions to the Program Design

During the pilot study, student orientation to the program was accomplished by a short computer tutorial which students completed before attempting the computer-based test. While this tutorial demonstrated the test features available to students, problems developed when the students were later unable to access this information during a test. The revised version of

the program included a new "on-line" help system that could be accessed by students throughout a given test.

The new help system was indicated by new icons that were included at the bottom of each screen. These icons operated to demonstrate the different features available in a test Three separate demonstrations illustrated the use of the navigational buttons, the elimination feature and the answering feature. Each of these demonstrations included the use of pop-up fields and animation techniques to convey their message to students. Test operation was suspended during the use of the help features.

Another option included in the revised design was a new review feature. This feature allowed students to immediately obtain their final score and to review their incorrect responses. This was accomplished through the use of a script that scanned tests searching for incorrect items. These were then presented with the incorrect answer and correct answer displayed.

Revisions to the Research Design

Information gathered during the pilot study was also helpful in revising the research plan for the main portion of the study. The following considerations evolved from noted inadequacies in the pilot study that ultimately led to problems in the comparison of test scores. The first consideration in the new research design was to provide sufficient motivation for students to maximally perform. Test items on the different

versions of the test would be controlled to the extent that each test consisted of the same items presented in the identical order. The study would include results on a series of tests rather than on one administration to ensure that the students were well versed in the use of the testing application. These and other considerations were also included in the final research design.

The Research Study

The major component of this research was to study student attitude and performance as it related to CAT. The comparisons made between computer administered and conventional paper-pencil tests involved an analysis of the relative student performances as measured by the multiple choice test scores for each format. Student attitudes were measured by pre and post surveys distributed at the beginning and end of the study. Finally, the usefulness of this computer administered test was evaluated in part through the computerized collection and analysis of test path data.

The study group involved fifty adult students from four classes of introductory computer studies at the Adult Basic level. The study took place at two community college campuses in British Columbia between June and December (1992). During the study, instruction was provided in a student-centred, self-paced format. The teaching situation included a microcomputer lab equipped with IBM XT and AT machines operating within the DOS 5.0 environment (Disk Operating System). The decision to use introductory computer classes as the study sample was based on two criteria. First, it was

believed that these students would have little prior experience with computers. Second, the use of computers in testing would be of additional pedagogical value to the students as it would familiarize them with another operating system (the Macintosh operating system was used for the testing).

The main portion of the study was conducted using an experimental method with two comparison groups (each alternating between the role of control and treatment group). Each of the four classes were randomly divided into two groups of roughly equivalent size. Each group was then administered a series of six multiple-choice tests during the course, with each test consisting of 20 multiple-choice items. These items were adapted from the instructors guide for use in this course (a report of the items used in the tests is included in appendix D). To account for possible differences in test difficulty the items on each version of the test were identical and presented in identical order.

For each test one comparison group consisted of the student scores obtained with pencil-paper tests while the other comparison group consisted of the student scores obtained with computerized tests. To account for possible differences in ability between the groups the following testing schedule was utilized: each group alternated between the computerized and written formats of the test (i.e., on the first test, group A was administered the written format while group B was administered the computerized format, on the following test the roles were reversed with group B taking the written test and group A the computerized test, etc.). At the end of the study each

student had taken six tests, three each in the written and computer formats. The test administration schedule is outlined in Table 1.

Table 1

Test Administration Schedule for the Study

Test Number	CAT format	Pencil-Paper Format
1	Group A	Group B
2	Group B	Group A
3	Group A	Group B
4	Group B	Group A
5	Group A	Group B
6	Group B	Group A

The resulting achievement scores were converted to mean scores for each of the test administrations and in addition grand means calculated for each format of the test. T-tests were then calculated and used to test the null hypothesis (i.e., that there were no significant differences in achievement scores between the two testing formats).

Before the test administrations students in both groups took part in a pre-survey by filling out a written questionnaire. The intent was to gauge students' prior attitudes about computers and computerized testing while collecting additional demographic information. Following the tests students were asked to take part in a written post-survey which asked similar questions but in addition collected students' impressions regarding design features included in the computer-based test (a copy of each survey is

included in Appendices E and G). The results of the surveys were tabulated for the purposes of comparison.

A script operating in the background of the CAT functioned to collect test path data for each student during tests three through six (a decision was made not to collect data during the first two tests so that students could become familiar with the format of the tests). The script recorded item order, item score, item response time, and total time for each test. The information was recorded each time a student accessed an item, therefore if a student changed an answer, multiple records were recorded for that item.

The test path data was written to word-processing files and later tabulated for comparison with similar data recorded for the written tests. This information included total time taken for a test and an item by item analysis of each test. In addition, the background script also recorded the frequency of use of the "review" and "elimination" features of the computerized test. This was done in order to facilitate a further analysis of those features in the software design through comparison with the student achievement scores on the computer-based tests.

Chapter Four

Research Findings

Pre -Survey Results

The pre-survey questionnaire distributed before the study yielded some informative demographic information about the study group as well as gauging student attitudes towards computers and computerized testing. Responses were grouped into categories and are summarized in Table 2.

Group Demographics

All students were enrolled in an introductory computer course at the Adult Basic Education level. The total sample consisted of fifty individuals, with thirty-three females and seventeen males (one individual withdrew from the study before the first test). The subjects ranged in age from nineteen to fifty seven years. The largest single grouping was aged twenty-five years or younger. On average, the students in the study sample were twenty-eight and a half years old and had been out of school for approximately eleven years.

The employment experience of those in the study varied but the majority of experiences were in the service sectors such as tourism and hospitality. Other types of experience included industry, labour, and office work. Experience in the health and education fields was also reported.

Student Responses Prior to Testing

Table 2

Student response summary for pre-survey questions

#1) gender	**male**	**female**				
	17	33				
#2) age of respondent	**(up to 25)**	**(26 to 30)**	**(31 to 35)**	**(36 to 40)**	**(over 40)**	**n/a**
	21	9	13	3	3	1
#3) yrs out of school	**(up to 5)**	**(6 to 10)**	**(11 to 15)**	**(16 to 20)**	**(over 20)**	
	15	11	12	7	5	
#4) work experience	**service**	**industry**	**labour**	**office**	**health**	**education**
	29	8	5	4	2	2
#5) do well on MC	**yes**	**no**	**n/a**			
	39	8	3			
#6) computer exp.	**none**	**(up to 2 yrs)**	**(3 to 5 yrs)**	**(over 5 yrs)**		
	33	10	6	1		
#7) previous test	**yes**	**no**				
	2	48				
#8) preferred format	**computer**	**written**	**n/a**			
	27	12	11			
#9) improve grades	**yes**	**no**	**n/a**			
	23	6	21			
#10a) advantage	**n/a**	**easier**	**convenient**	**faster**		
	17	17	10	10		
#10b) disadvantage	**n/a**	**confusing**	**intimidate**	**eye fatigue**		
	40	8	5	2		

NOTE. Advantages/Disadvantages sum inconsistently as frequencies of less than two are not shown. (N=50)

The majority of students surveyed indicated that they generally did well on MC type tests. Two individuals had previously undertaken such a test on a computer. Most respondents had no experience with computers before the study (66 percent) with an additional 20 percent of respondents having less than two years experience.

When asked which format of the test they would prefer, about half believed that they would prefer the computer format (54 percent), with the remainder split evenly between preferring a written format (24 percent) and no preference (22 percent). Additionally, 46 percent of students indicated that they believed that computerized testing might improve their overall grades while 12 per cent believed that they would not. Forty-two per cent did not respond to this item.

When students were asked what they expected the advantages of computerized testing might be, many did not respond (34 percent). However, an equal number of students indicated that they expected the computer tests to be easier than the written tests. Other expectations were that computer tests would be faster and more convenient. When asked what they expected the disadvantages of computerized testing might be, the majority did not respond (80 percent), while a few expressed concerns that the tests might be more confusing or intimidating.

The general mood regarding computerized testing was a positive one. However, many students expressed some uncertainty about the potential results. A complete matrix of student responses to the Pre-survey and a sample of the Pre-survey questionnaire is provided in Appendices E and F. These findings are summarized in Table 2.

Post-Survey Results

The post-survey questionnaire distributed after the study gauged the changes in student attitudes regarding computerized testing. The survey was administered following their exposure to this type of test. In addition, valuable feedback about the test's software and interface design was obtained through the use of this survey. The sample size for this portion of the study was reduced to 43 from 50 (one student withdrew from the study and in addition 6 Post-survey questionnaires were not returned). However, the demographics of the smaller sample were sufficiently similar to facilitate comparison. As in the pre-survey, student responses were grouped into categories. The results of this analysis are summarized in Table 3.

Student Responses After Testing

Following the testing schedule, the majority of students indicated a preference for the computerized format (72 percent), over the written format (9 percent). In addition, 19 percent of students were unable to choose a preferred format. The results here compare favourably with those obtained in the pre-survey with clear increases shown in the number of respondents favouring computerized tests and a corresponding decrease in those preferring the written format. The number of respondents choosing no answer remained relatively unchanged.

Table 3

Student response summary for post-survey questions

#1) gender	**male**	**female**				
	13	30				
#2) age of respondent	**(up to 25)**	**(26 to 30)**	**(31 to 35)**	**(36 to 40)**	**(over 40)**	**n/a**
	17	10	10	2	3	1
#3) yrs out of school	**(up to 5)**	**(6 to 10)**	**(11 to 15)**	**(16 to 20)**	**(over 20)**	**n/a**
	13	8	8	7	3	4
#4) preferred format	**computer**	**written**	**n/a**			
	31	4	8			
#5) improve grades	**yes**	**no**	**n/a**			
	16	21	6			
#6) used elimination	**yes**	**no**	**n/a**			
	30	5	8			
#7) review answers	**yes**	**no**	**n/a**			
	29	6	8			
#8) used review	**yes**	**no**	**n/a**			
	31	4	8			
#9) advantage	**scoring**	**easier**	**faster**	**elimination**	**fun**	
	12	11	8	8	5	
#10) disadvantage	**n/a**	**unfamiliar**	**waiting**	**too rushed**		
	29	5	2	2		

NOTE. Advantages/Disadvantages sum inconsistently as frequencies of less than 2 are not shown. (N=43)

In contrast to this, when students were again asked whether they believed that the computer tests would improve their grades, only 37 per cent answered yes while 49 per cent responded no. In addition, only 14 per cent of respondents gave no answer. The results show that while students generally expressed a clear preference for the computerized format of testing, they were now unsure whether they could improve overall grades.

When asked to list the advantages of computerized testing, students were more confident in their assessment with all respondents listing at least one advantage, while 34 per cent responded with no answer in the pre-survey. The most frequently listed advantage was the autoscoring feature of the computer test. An equivalent number of students indicated that the computer tests were easier to use than written tests, thereby conforming with earlier student responses obtained in the pre-survey. Other frequently listed advantages of the computer format included the use of the elimination feature and that the tests were generally faster and more fun.

When asked to list the disadvantages of computerized testing, the majority of students still responded with no answer (67 per cent) while the total frequency of disadvantages amounted to only 20 percent. The most frequently listed disadvantage was an unfamiliarity with computers. While expressed relatively infrequently, the final two disadvantages listed relate to the frequent waits experienced by students taking the computerized tests.

Overall, the analysis and comparison of Post-survey and Pre-survey responses indicated that students generally held positive attitudes towards computerized testing and that this positive view increased with their use of the computer administered tests. Students clearly indicated a preference for the computer format and generally listed more advantages than disadvantages in its use. A complete matrix of student responses to the Post-survey and a sample of the Post-survey questionnaire is provided in Appendices G and H. These results are summarized in Table 3.

Reported Use of Computer Test Features

The use of a post-survey also allowed students to report on their use of the newly designed test features provided in the computer format of the test. These features included the ability to eliminate distracters from view during a test and the ability to review questions and answers before the computer giving a final score for a test. Finally, students had the option to review their incorrect responses after completing and marking a test.

The elimination feature was designed to help students arrive at an answer by allowing the removal of distracters from the field of view. This feature is the rough equivalent of drawing lines through a choice on a written test although this action could be reversed on the computer tests. When asked about their use of the elimination feature, 70 per cent of the students reported using it while an additional 12 per cent reported not using it. The remaining 19 per cent responded with no answer to this item.

The review feature allowed students to browse through the questions before having the test auto-scored. When asked about their use of this feature 60 per cent of respondents reported using it while 14 per cent reported not using the feature. Again, 19 per cent responded with no answer. Similar results were also reported for student use of the post-test review feature with 72 per cent of respondents reporting usage.

Comparative Analysis of Test Scores

Test scores on the test administrations were considered in two groups with differing sample sizes. The first two sets of scores were considered separately to gauge the potential differences in scores obtained on the initial administration of the computerized test. The remaining four test administrations were subjected to more intense analysis although for practical purposes the sample size for this comparison was reduced. One class did not have access to an additional computer that operated outside the confines of the network and this class was therefore removed from the study. This decision helped to facilitate the collection of the test path data as the remaining classes had access to the computer provided for this purpose.

In the case of the first two tests the study sample consisted of 49 individuals writing two different tests (one person withdrew from the study after the pre-survey). The administration of the two tests therefore resulted in of a total of 98 test scores evenly distributed between the written and computer formats. The results were obtained by the auto-scoring of the computerized tests and the manual marking of the written tests.

The results were compared in two ways: Mean scores were calculated for each group (A and B) with raw test scores on both the computer and written tests used in the calculation (maximum score was twenty). The resulting mean scores were compared using a two-tailed t-test (with alpha

.05 and 96 df) to determine if there were differences in the scores between groups regardless of testing format. The results indicated that there was no significant difference between these groups.

After ruling out differences in achievement between the two groups, mean scores were calculated for each test administration and grand means were calculated for each format of the test. Similarly, raw test scores were used for each format of the test with scores from each group contributing to the mean scores (maximum score was again twenty). A two-tailed t-test (with alpha of .05 and 96 df) was conducted to determine if there were differences in the scores obtained using the different testing formats. Considering that this was the first use of the computerized test, these scores were considered separately as a difference in scores (favouring the written format), was expected due to students' unfamiliarity with the computerized format of the tests. However, the results indicated clearly that there was no significant difference in achievement using the different test formats. The calculated statistics for tests one and two are summarized in Table 4.

Table 4

Overall statistics for scores in test administrations one and two

Group	n	grand mean	variance	pooled	t score	Critical t
A	48	14.44	9.06	10.55	0.43	1.66
B	50	14.16	11.97			
Computer	49	14.24	9.40	10.59	0.17	1.66
Written	49	14.35	11.73			

Note: Maximum test score = 20

For the remaining four test administrations, a similar analysis was applied to the reduced sample size. The study group consisted of thirty-eight students from the original sample (one class was removed from the study due to problems with access to the needed equipment). During this part of the study, each student completed a total of four tests. The administration of the tests therefore resulted in a total of 131 test scores evenly distributed between the written and computer formats (with the exception of test number six where two individuals failed to complete the computer version of the test). As in the first two test administrations, the results were obtained by a combination of auto-scoring and manual marking. The calculated statistics for these tests are summarized in Table 5.

Table 5

Overall statistics for scores in test administrations three to six

Group	n	grand mean	variance	pooled	t score	Critical t
A	76	15.05	5.33	3.19	3.42	1.96
B	74	14.05	1.00			
Computer	74	14.59	3.00	2.66	0.23	1.96
Written	76	14.53	2.33			

Note: Maximum test score = 20

The results of the last four tests were compared in two ways: First, mean scores were re-calculated for each group (A and B) and a two-tailed t-test (with alpha .05 and 148 df) was conducted to determine if there were differences in achievement between the two smaller groups.

Initial analysis of the individual means showed that group B scored on average one point lower than group A on each test administration. The results of the t-test indicated that these differences in achievement between the two groups were significant. Over the four test administrations, group B scored consistently lower than group A.

Noting the observed difference in achievement between the two groups, mean scores were then calculated for each test administration and grand means were calculated for each format of the test. A two-tailed t-test (with alpha of .05 and 148 df) was again conducted to determine if there were differences in achievement attributable to the different testing formats. The results supported the data gained in the first two test administrations indicating that there was no significant difference in achievement using the computer testing format. A detailed summary of test scores for all six test administrations is included in appendix C.

Following this finding, further comparisons were conducted on a test by test basis to determine whether there were significant differences in achievement on the individual test administrations. This was accomplished with a series of two-tailed t-tests (with alpha of .05).

The test by test analysis supported the overall findings by indicating that for each of the six test administrations there were no significant differences in achievement between the two formats. Summary statistics for this test by test comparison are provided in Table 6.

Table 6

Individual test statistics for test administrations one to six

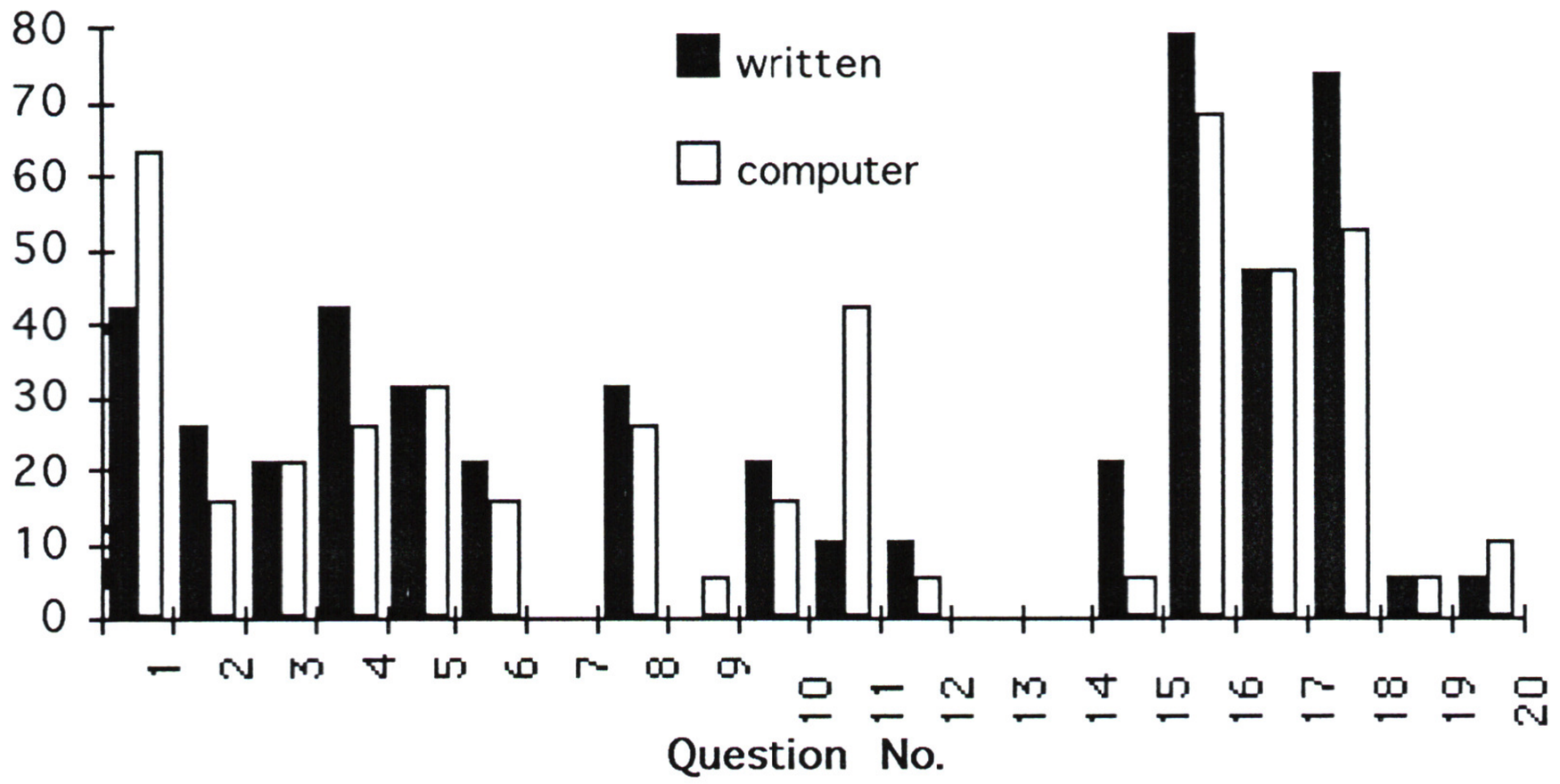
Test No.	format	n	mean	variance	pooled	t score	critical t
1	computer	24	13.96	7.52	10.19	0.18	2.01
	written	25	13.80	12.75			
2	computer	25	14.52	11.43	10.98	0.42	2.01
	written	24	14.92	10.51			
3	computer	19	15.68	3.78	5.25	0.85	2.03
	written	19	15.05	6.72			
4	computer	19	11.95	7.50	5.52	1.23	2.03
	written	19	12.89	3.54			
5	computer	19	15.58	6.15	6.62	1.76	2.03
	written	19	14.11	7.10			
6	computer	17	15.24	5.82	6.06	0.99	2.04
	written	19	16.05	6.27			

Note: Maximum test score = 20

Item by Item Analysis of Test Scores

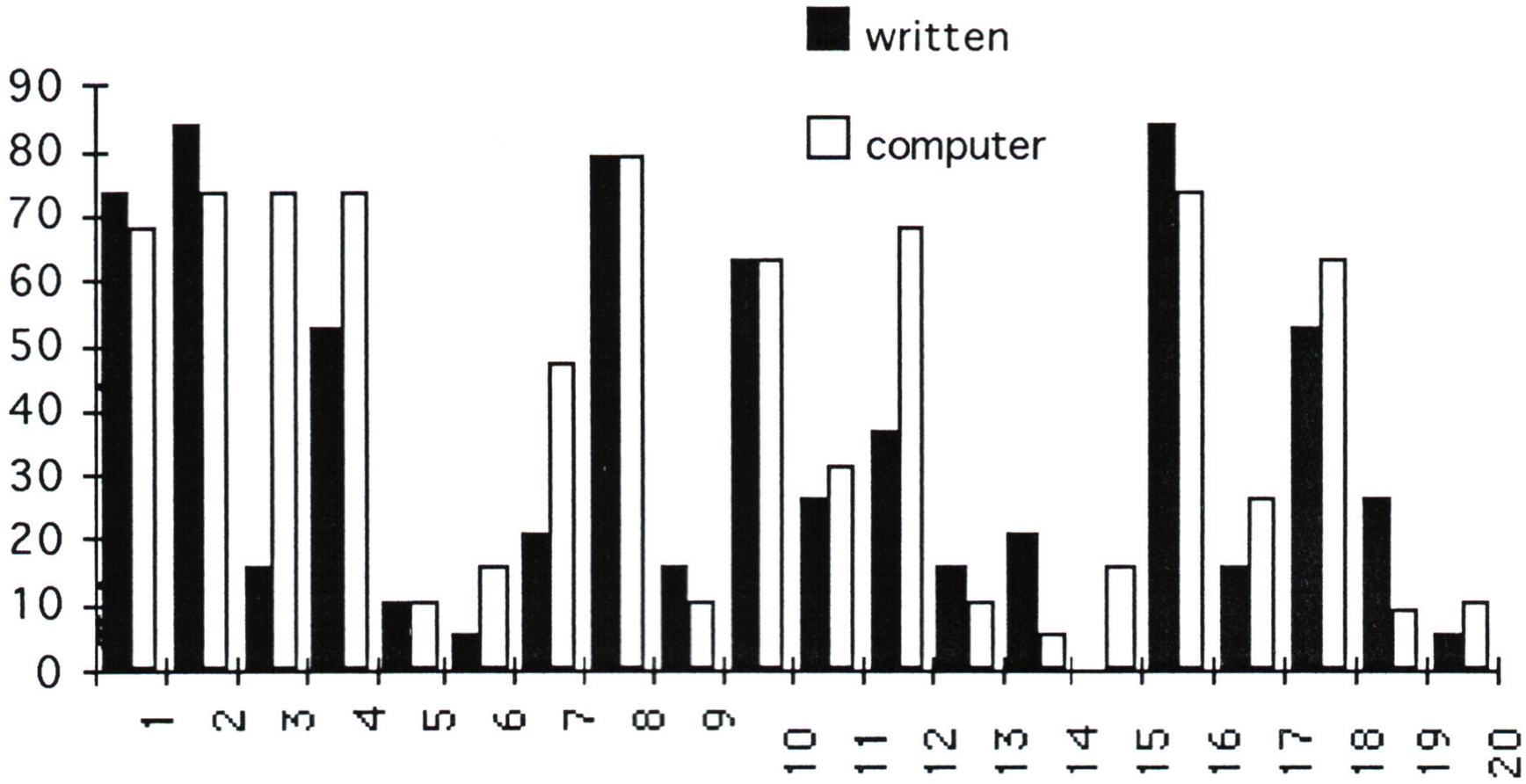
An item by item analysis was conducted for tests three to six with the number of incorrect responses on each test item and each format recorded. The results were graphed and are presented in Figures 1 to 4. Overall, the graphical representation of test response patterns gives an indication of the comparative difficulty of individual items on a test (the maximum values on a graph indicating more difficult items). The pattern of maximum and minimum values is consistent for both administrations of a given test. This trend was shown throughout the successive test administrations and shows that difficult items on the computer format of a test corresponded with those of the written format and vice versa.

Figure 1 - TEST 3 INCORRECT RESPONSE PATTERNS FOR DIFFERENT TEST FORMATS



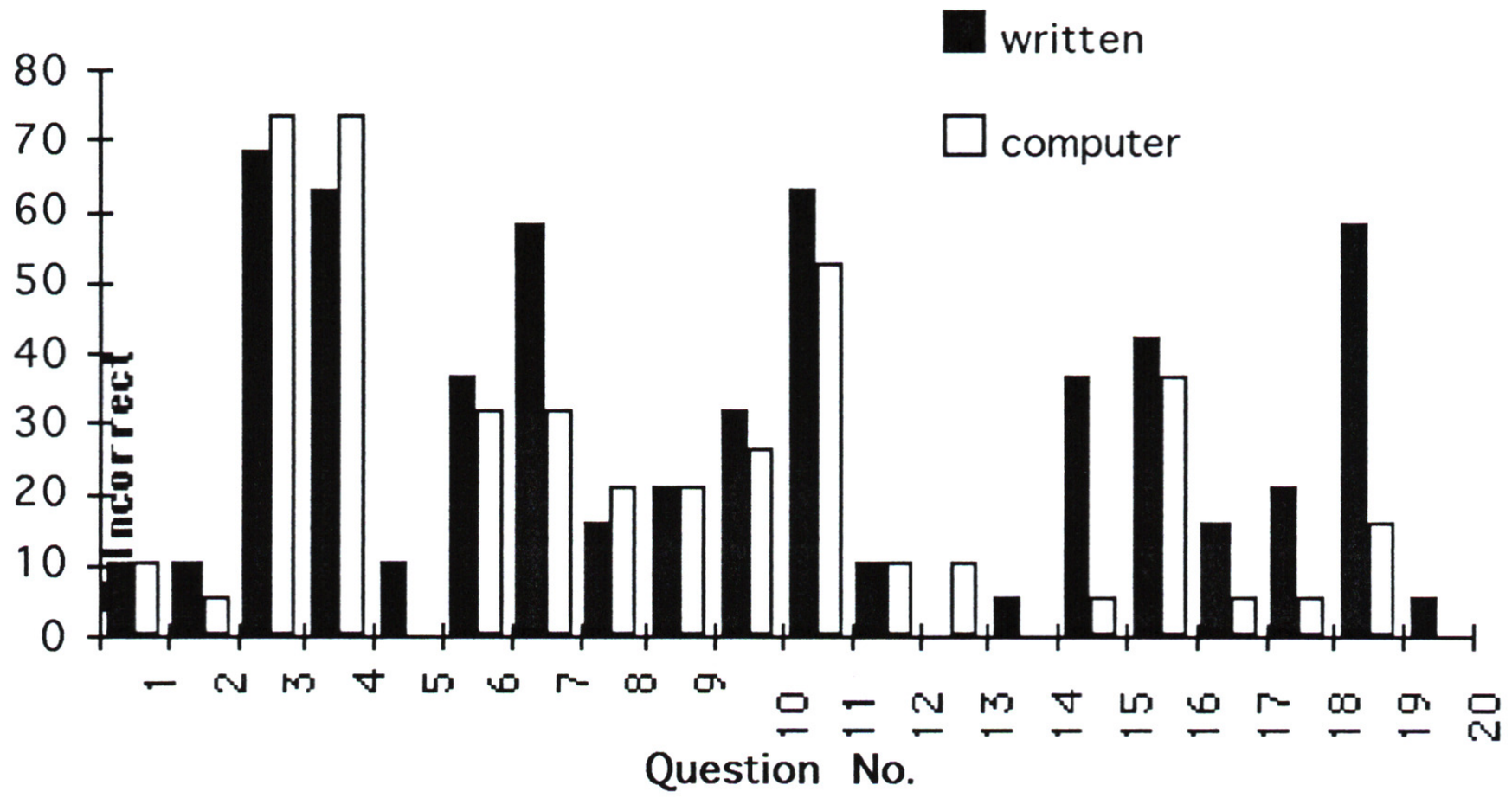
N=38

Figure 2 - TEST 4 INCORRECT RESPONSE PATTERNS FOR DIFFERENT TEST FORMATS



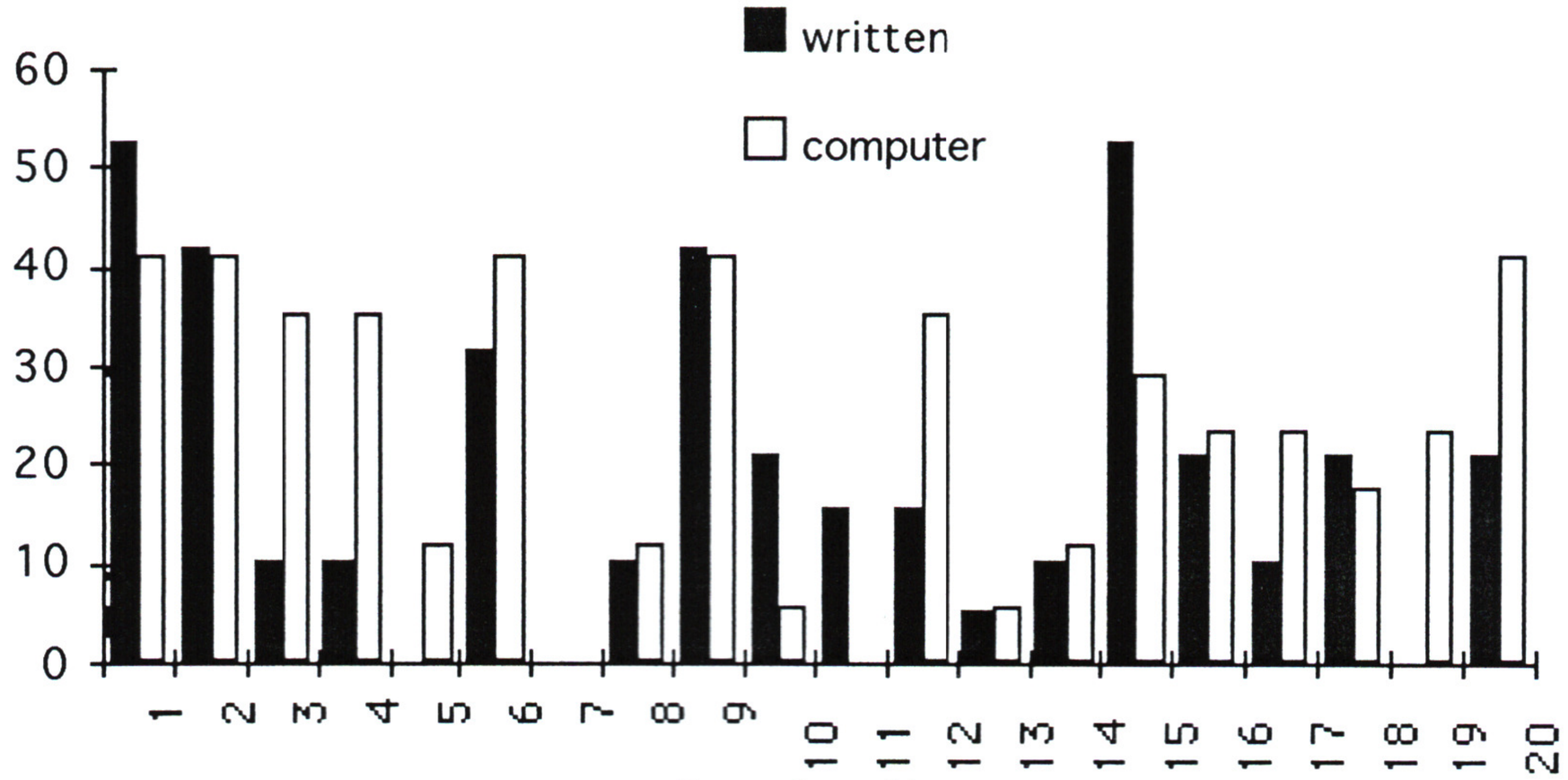
N=38

Figure 3 - TEST 5 INCORRECT RESPONSE PATTERNS FOR DIFFERENT TEST FORMATS



N=38

Figure 4 - TEST 6 INCORRECT RESPONSE PATTERNS FOR DIFFERENT TEST FORMATS



N=36

The graphical representation of the response patterns did illustrate some differences in the magnitude of difficulty on some test items between the different formats. For example in Test three the written versions of the test showed a greater number of errors for items four and eighteen while the computer version showed a greater number of errors for item number eleven. While some variation would be expected, this variation between test formats was demonstrated on all four test administrations.

Analysis of Test Path Data

The provision of a number of sub-programs imbedded in the design of the computer administered testing program allowed for the collection and analysis of test path data. The information recorded by the computer included test score, test administration time, and frequency of use of the test features. The latter was accomplished by analysis of a student's test path through individual tests.

Comparison of Test Administration Times

In addition to the qualitative and quantitative analysis undertaken for tests three through six, the collection of test path data allowed for a comparison of test administration times as the time taken for each written test was also recorded with the use of a stopwatch. The resulting test times were compared in two ways.

Mean test times were calculated for each group and a two-tailed t-test (with an alpha of .05 and 148 df) was conducted to determine if there were differences between the two groups. The results of the test indicated that group B took significantly more time to complete the tests than did group A.

Initial analysis of the mean test administration times showed that for the most part, the computer format of the test required more time than did the written tests. For example in test three the computer version of the test took on average two minutes longer than the written version. Test four showed the largest difference with the computer test taking about five minutes longer. A summary of test administration times is included in appendix C.

One exception to this trend was test five where the computer version required approximately one minute less time than did the written test. In general, the differences in test times were less pronounced on the later tests with test six exhibiting a difference of about one and a half minutes between the two different formats (with the written tests requiring less time).

Following the initial analysis of the test means, grand means were then calculated for each format of the test and a two-tailed t-test (with alpha of .05 and 148 df) was again conducted to determine if the observed differences in test administration times were significant. The results clearly indicated that there was a significant difference in the time taken to complete the tests with the computer format requiring more time on average than did

the written tests. A complete summary of the calculated statistics for the recorded test times is provided in Table 7.

Table 7

Combined test times and statistics for test administrations three to six

Group	n	grand mean	variance	pooled	t score	Critical t
A	76	15.1	6452	32363.76	2.99	1.96
B	74	16.5	58188			
Computer	74	16.8	35202	27816.49	4.41	1.96
Written	76	14.8	19893			

Note: All test times were recorded in seconds but have been presented in minutes.

Frequency of Use of the Test Features

Analysis of the test path data allowed further feedback regarding the frequency of use of the included test design features. Analysis of the individual test paths showed a more objective measure of the frequency of use of each design feature and allowed for comparison with the frequencies reported by students on the post-survey questionnaire.

On each administration of the computerized tests studied there were nineteen participants with the exception of the two missing scores from test number six. Therefore, a total of seventy-four test paths were recorded. Overall, the number of students using the design features was fairly high at about 76 per cent of the total usage. This figure was somewhat higher than the estimate of 70 per cent estimated by the post-survey responses. This

total usage broke down to include a 62 per cent usage of the elimination feature, and roughly 83 per cent usage for both of the test review features. The results of this analysis are included in Table 8.

Table 8

Frequency of use of computer test design features

Test Number	Elimination Feature	Review Questions	Review Scored Test	Sample Size
Three	11	17	16	19
Four	19	19	18	19
Five	10	12	14	19
Six	6	13	15	17
Total	**46**	**61**	**63**	**74**

Finally, the use of the measured test design features was compared with regards to individual achievement on the computerized tests. In a rudimentary analysis the ten highest and ten lowest test scores were compared in terms of general use of the test design features.

Analysis of the higher scores showed that nine out of ten scores had been obtained using the elimination feature (90 percent) and 8 out of 10 scores (80 percent) had involved the use of the review feature before scoring. In comparison, analysis of the lower scores showed that 3 out of ten scores (30 percent) were obtained using the elimination feature and 6 out of ten scores (60 percent) had involved the use of the review feature.

Chapter Five

Summary and Discussion

Summary of Purpose and Methodology

The primary goal of this research was to analyze the validity of claims made by some educators who believe the use of computer administered tests were in some way disadvantageous to students, (Moreno, 1985; & Spray, et al., 1989). While such claims have some support in the literature, many researchers believed that these problems were relatively minor and were balanced by the potential administrative benefits realized through the use of such systems. Analysis of the problem was addressed through a number of research strategies undertaken in this study.

First, the issue of documented inequivalencies in test results between written and computerized formats was considered a symptom of poor interface and software design and a resulting increase in test anxiety. The objective was to design a better test that could reduce the potential for this anxiety and thereby reduce the observed differences in test scores.

Development of the software was on-going and intensive with the process taking approximately 12 months before the main part of the research began. Part of this process was the extensive consultation with students and teachers. The resulting program was pilot tested with student input leading

to the inclusion of several new test features in the application and to additional changes in the user interface.

Following the development of the testing application, the next objective was to gauge its effect on student beliefs about computerized testing. This was accomplished by survey questionnaires that were distributed to students before and after the main portion of the study. The comparative analysis of these surveys yielded some subjective information about the effects of the test while also collecting some useful demographic information about the study group. The post-survey also allowed for the collecting of student feedback regarding the new design features included in the test.

The central objective then was to undertake a comparative analysis of test scores obtained using written tests and computer tests administered by the new application. This was accomplished by an experimental design that saw the study sample randomly divided into two equal groups. The two groups then alternated through six different test administrations with each group completing three tests in the written format and three tests on the computer. Each test consisted of twenty multiple choice items presented in the same order on either format. Scores were calculated by computer in the case of the computerized tests and by the teacher for the written tests. Following the collection of the test scores, statistical tests were performed to determine if there were significant differences in achievement between the two groups and between the two formats of the test.

In addition to the quantitative analysis described above, a further qualitative analysis was conducted around the different formats of test administration. An item by item analysis was conducted for each test administration for both formats of the test. A test item response pattern was illustrated by plotting the number of incorrect responses for each test item. These response patterns were then used to gauge the comparative difficulty of each test item as they appeared on either format of the tests.

A further goal was to collect objective information about the use of the computerized tests. This was accomplished through the development of a number of sub-routines operating in the background of the testing application that recorded the action of students as they proceeded through a test on the computer. The analysis of the resulting test path data allowed a comparison of administration times between the two formats of the tests and provided further feedback regarding the use of the included design features.

The objectives described here were then condensed into the form of three hypotheses that formed the central purpose of this study. The hypotheses were tested using the above noted combination of surveys, test scores, test administration times, and the computer test path data.

Summary of Hypotheses

1) The first hypothesis was that a suitable computer administered test could be designed (in consultation with students and teachers) that would elicit a more positive student attitude towards computerized administered forms of testing. This hypothesis was tested through the comparison of student questionnaires administered before and after the use of the program.

2) The second hypothesis was that the computerized test would demonstrate that the test scores obtained on computerized tests were not significantly different from scores obtained on paper and pencil tests (the null hypothesis). This hypothesis was tested using a statistical comparison of test scores and a qualitative analysis of response patterns.

3) The third and final hypothesis was that the computerized format of a test would allow for significant increases in test efficiency as measured by the time taken to write a given test, (i.e., that computer tests would require less time to complete). This hypothesis was tested by a statistical comparison of the recorded test times.

Summary of Findings

The results of the pre survey showed that students believed they would prefer the computer administered format of the test. Before testing approximately twice as many students believed that they would prefer the computer format over the written format of a test (54 percent versus 24 percent). At the same time, many students were unsure whether they would prefer either version of a test (22 percent). When asked to list the advantages of computerized testing many responded with no answer, (34 percent), while an equal number expected them to be easier than written tests.

Results of the post-survey showed clearly that students did indeed prefer the computerized format of the tests. Following testing eight times as many students now preferred the computer format over the written format of a test (72 versus 9 percent). Clearly, the student's exposure to the computer administered tests caused a further shift in student attitude in favour of computerized testing. Additionally, when asked to list the advantages of the computer format, all students were able to list at least one advantage. In comparison, when asked to list the disadvantages of computerized tests, 67 percent of the students failed to report even one.

The completed questionnaires point to success in the revisions undertaken in software and interface design. The results of the comparisons

support the premise that good design could influence student attitude regarding computer administered tests. It is important to note that the use of the computer tests was in itself a novelty for the majority of students and this may also have affected student attitudes.

Summary of Test Scores

The quantitative analysis of test scores also conclusively supported the premise that students were not being disadvantaged by taking computer administered tests. Initially t-tests were conducted between the overall means of either comparison group (A and B) to determine if there were between group differences regardless of format. The results indicated that on the initial test administrations there were no significant differences. Following this comparison, t-tests were again conducted between the overall mean scores calculated for each format of the test and again these yielded no significant differences between the different formats.

For the later administered tests (three through six) the comparison of scores also yielded no significant differences in test score attributable to format. This was true despite the noted differences attributed between groups regardless of the test format. Effectively, the counterbalanced nature of the experimental design (alternating test administrations between the two comparison groups) compensated for the inherent differences in ability between the comparison groups. More importantly, the findings demonstrate that even on the initial administrations of a test, students scored similarly on

the computer tests, despite their inexperience with the format or the testing program. The results of this statistical analysis clearly support the second hypothesis of this study.

A detailed item by item analysis of the last four test administrations revealed similar patterns of response on both formats of the tests. Items which many students found difficult on the computer format of a test also tended to be the more difficult items on the written format and vice versa. While the overall similarities in the pattern of student responses was closely related throughout all four tests, many items initially appeared to be more difficult on one version of a test than the other. Many of these differences can be attributed to the differences in ability observed between the two comparison groups. For example when the lower achieving students (group B) were writing the computer format of a test the computer format would show a greater percentage of incorrect responses on some items. Similarly when the higher achieving students were writing the computer format of a test the same items would appear to show a lower percentage of incorrect responses than those in the written version of a test.

Despite these relatively small differences the overall pattern of test responses remained similar across both formats of the tests. This qualitative analysis of response patterns when coupled with the results of the statistical comparisons lends further support that the computer and written formats of the tests were equivalent in terms of achievement.

Summary of Test Administration Times

The hypothesis made regarding the efficiency of computerized tests (in terms of administration times) was not supported by this research. Comparisons made between the test administration times obtained on the computer and written formats were summarized in the form of mean scores for each administration of the last four tests. The results showed that on each test the written format of a test took less time to complete than did the computer format of the test, (the exception to this was test number five where the computer format took less time to complete). Grand means for each format of the test over the last four tests were calculated. A t-test revealed that there were significant differences in test administration times between the two formats of the tests, with the computer format requiring more time on average than the written format.

In retrospect, it does seem more reasonable to assume that students would require more time in the use of a new or unfamiliar program as was the case for the newly developed computer administered test. This would explain why the computer format of the tests required significantly more time. If the observed difference could be attributed to this unfamiliarity, the differences might be expected to disappear after successive test administrations. Further analysis of the test administration times does reveal a trend. The largest differences in tests times were observed on tests number three and four. On test number six, these differences were greatly reduced and on test number five the computer format actually required less time.

This information supports the premise that it was the students' relative unfamiliarity with the computer format of the tests that contributed to the significant differences observed in the administration times. Another important note is that the average difference in the test administration times amounted to only two minutes.

Summary of Test Path Data

The collection of test path data by the computer allowed the collection of more objective information about the use of the computer administered testing program. Specifically, it provided the frequency of usage of all of the included test design features. Overall, the actual usage of the design features was relatively high at 76 per cent, (somewhat higher than that reported in the post-survey questionaire). The most popular features were the review features which allowed students to review their responses before scoring a test and secondly to review their incorrect responses after the test was scored. Inclusion of these and other test features, (such as the elimination feature) allowed a similar range of test-taking behaviours as possible in written formats of multiple choice testing. The reported frequencies indicate that these were necessary and useful features contributing to the overall success of the program design. Use of these features may also have contributed to the success of some students as rudimentary analysis of tests scores showed that use of these features (particularly the elimination feature) may have been linked to higher achievement scores.

Conclusions

In conclusion, a number of the noted concerns of educators regarding computer administered testing have been addressed by this study. Through careful application of software development guidelines and through consultation with teachers and students a computer administered test was developed. The success of the program was linked to its ability to allow a similar range of test taking behaviours as those allowed by the written format. When students were asked to indicate a subjective preference for a given format of the test they clearly indicated a preference for the computerized version. Further, the strength of this preference seemed to increase with their usage of the tests.

The combination of the survey findings and the statistical results would seem to indicate that students were clearly not disadvantaged by taking their tests using this computer administered testing application. These findings might be attributed to the inclusion of several new design features in the computer administered test. Their inclusion may have addressed previous inequities between the computer and written formats of multiple choice tests. However, the use of computerized tests may require more time for their administration. As indicated, the computer administered tests in this study required on average two minutes longer to complete than the written tests. However, this additional requirement may be lessened as students become more familiar with the format of these computer tests.

Recommendations from this Research

The findings of this research have implications regarding the design and use of computer administered testing programs within the instructional setting where it was implemented and tested. However, because the study group consisted of only a small sample of adult students, the results may not be generalizable to other student groups or to other academic areas.

Factors which might have contributed to a more reliable study in the given setting would include a larger sample group more evenly distributed by gender, (most individuals in the study group were female). This would have allowed a more thorough analysis of the test scores on each individual test administration and would also have allowed for gender comparisons between achievement scores obtained with the two different formats.

Additionally, a longer sequence of testing using more reliable instruments would have further strengthened the study. Comparisons in the mean scores obtained in further test administrations (i.e., greater than six) might also have indicated clearer trends in achievement and test times as the students would become more familiar with the new format. Finally, a clearer link could have been established between the use of the test features and achievement on these tests if the test path data had been collected differently. Information such as this would provide researchers with more complete information regarding the long term benefits of computer administered testing.

Recommendations for Future Research

Further studies are still needed to confirm the findings of this research. These studies could include the development of further software design features to be included in future computer administered tests. Further research could document the effective benefit of such features for students and teachers. Another area which could benefit from further research would be the replication of studies such as these within other educational contexts. These could include highschool level science students or perhaps students in other subject areas where multiple choice testing is commonly used, (i.e. social sciences, geography and psychology). This type of information could be used to demonstrate the versatility of the computer testing format.

Finally, the use of computer based programs to aid in educational research has been demonstrated in this study through the use of test path data collected by the computer. This use of Hypercard scripts was a powerful and objective tool in collecting data about student performance. Further, because the script operated in the background of the computerized test, data collection was unobtrusive and did not interfere with student behaviours. For these reasons the use of computers in educational research should be encouraged. Overall, the proper use of computer hardware and software may prove to be a powerful tool in the hands of all educators. Further research is needed to demonstrate its potential benefits and disadvantages.

References

Alessi, S., & Trollip, S. (1985). *Computer Based Instruction, Methods and Development.* Englewood Cliffs, NJ:Prentice Hall.

Anderson, P., & Schoner, J. (1987). Testing: 1,2..5,2,3..641..999 -Testing The MDT Technique Applied to Science Education. *Report published by Eric Clearinghouse*. (ED318789).

Barer-Stein, T. & Draper, J. (1989). *The Craft of Teaching Adults.* Toronto, Canada:Culture Concepts Inc.

Borg, Walter, & Gall, Meridith D. (1989). *Educational Research* (5th ed.). New York, NY:Longman.

Bowers, D., & Tsai, C. (1990). Hypercard in Educational Research: An Introduction and Case Study. *Educational Technology* , 2, 19-24.

Brookfield, Stephen D. (1983) *Adult learners, adult education and the community* . Milton Keynes, England: Open University Press.

Brookfield, Stephen D. (1986) *Understanding and facilitating adult learning: a comprehensive analysis of principles and effective practices.* Milton Keynes, England: Open University Press.

Burns, Edward (1990). Multiple Choice - Computer Diagnostic Test Designs. *Educational Technology* , 1, 49-53.

Collins, M., & Fletcher, P.(1985). Using Computer Tests to Identify Areas for Remedial Teaching of Biology. *Journal of Computers in Math and Science Teaching* , 2, 36-37.

Darkenwald, G. & Merriam, S. (1982). *Adult Education: Foundations of Practice*. New York, NY:Harper and Row.

Feuer, D. (1986). Computerized Testing: A Revolution in the Making. *Training* , 4, 80-86.

Fletcher, P., & Collins, M. (1986). Computer Administered Vs. Written Tests -Advantages and Disadvantages.*Journal of Computers in Math and Science Teaching* , 4, 38-43.

Frederico, P., & Liggett, N. (1989). Computer-Based and Paper-Based Measurement of Semantic Knowledge. *Report from the Navy Personnel Research and Development Center*. San Diego, CA.

Graesser, A. (1992). Design of Optimal Person-Artifact Interfaces: Contributions from Cognitive Science.*Education andTrainingTechnology International* , 29(1),15-21.

Howell, David C. (1989). *Fundamental Statistics for the Behavioral Sciences* (second ed.). Boston, MA:PWS-Kent Publishing Company.

Jackson, B. (1988). Comparison between computer-based and traditional assessment tests and their effects on pupil learning and scoring. *Science Educational Notes*, Englewood Cliffs, NJ:Prentice Hall.

Kapes, J., & Vansickle, T. (1992). Comparing Paper-Pencil and Computer-Based Versions of the Harrington-O'Shea Career Decision-Making System.*Measurement and Evaluation in Counselling and Development* , 25, 5-13.

Knowles, M. (1970). *The Modern Practice of Adult Education*. New York,NY:Association Press.

Knox, A. (1981). *Adult Development and Learning*. San Francisco, CA:Jossey Bass.

Knox, A. (1986). *Helping Adults Learn*. San Francisco CA:Jossey Bass.

Mizokawa, D. (1991). Computer Managed Testing in Schools. *Educational Technology* , 1, 21-25.

Mizokawa, D., & Hamlin, M. (1984). Guidelines for Computer Managed Testing. *Educational Technology* , 12, 12-18.

Moore, G.A.B., & Waldron,M.W. (1981). *Helping Adults Learn.*. Guelph:University of Guelph, Office for Educational Practice.

Moreno, K.(1985). Effects of computerized tests versus paper and pencil test administration on examinee performance. *Report from the* Navy *Personnel Research and Development Center*. San Diego, CA.

Olsen, J. (1990). Applying Computerized Adaptive Testing in Schools. *Measurement and Evaluation in Counselling and Development*, 23,31-38.

Percival, F., & Ellington, H.(1984). *A Handbook of Educational Technology*. New York, NY:Nichols Publishing Company.

Rubenson, K. (1983). Barriers to Participation in Adult Education. *A background paper prepared for the Task Force on Skill Development Leave and the Department of Employment and Immigration*. Vancouver:University of British Columbia.

Province of British Columbia (1990). *The Graduation Program Response Draft* . Victoria, British Columbia:Ministry of Education.

Sarti, Luigi (1992). Human/Computer Interfacing in Educational Environments. *Education andTrainingTechnology International*, 29(1), 52-62

Schuldberg, David (1988). Indices of individuals' sensitivities to computerized test administration. *Report published by Eric Clearinghouse*., (ED306294).

Smith, Richard L. (1987). Computer Assisted Testing - a Bibliography. *Journal of Computers in Math and Science Teaching*, 1, 73-74.

Spray, J., Ackerman, T., Reckase, M. & Carlson, J., (1989). Effect of Medium of Item Presentation on Examinee Performance and Item Characteristics. *Journal of Educational Measurement*, 26, 261-271.

Vansickle, T.(1992). Test-Retest Equivalency of the Computer-Based and Paper-Pencil Versions of the Strong-Campbell Interest Inventory. *Measuring and Evaluation in Counselling and Development* , 22, 88-93.

Waugh, M. (1985). The Effects of Microcomputer-Administered Diagnostic Testing on Immediate and Continuing Science Achievement and Attitudes. *Journal of Research in Science Teaching*, 22(9), 793-805.

Wise, S., & Plake, B. (1990). Computer-Based Testing in Higher Education. *Measurement and Evaluation in Counselling and Development*, 23, 3-10.

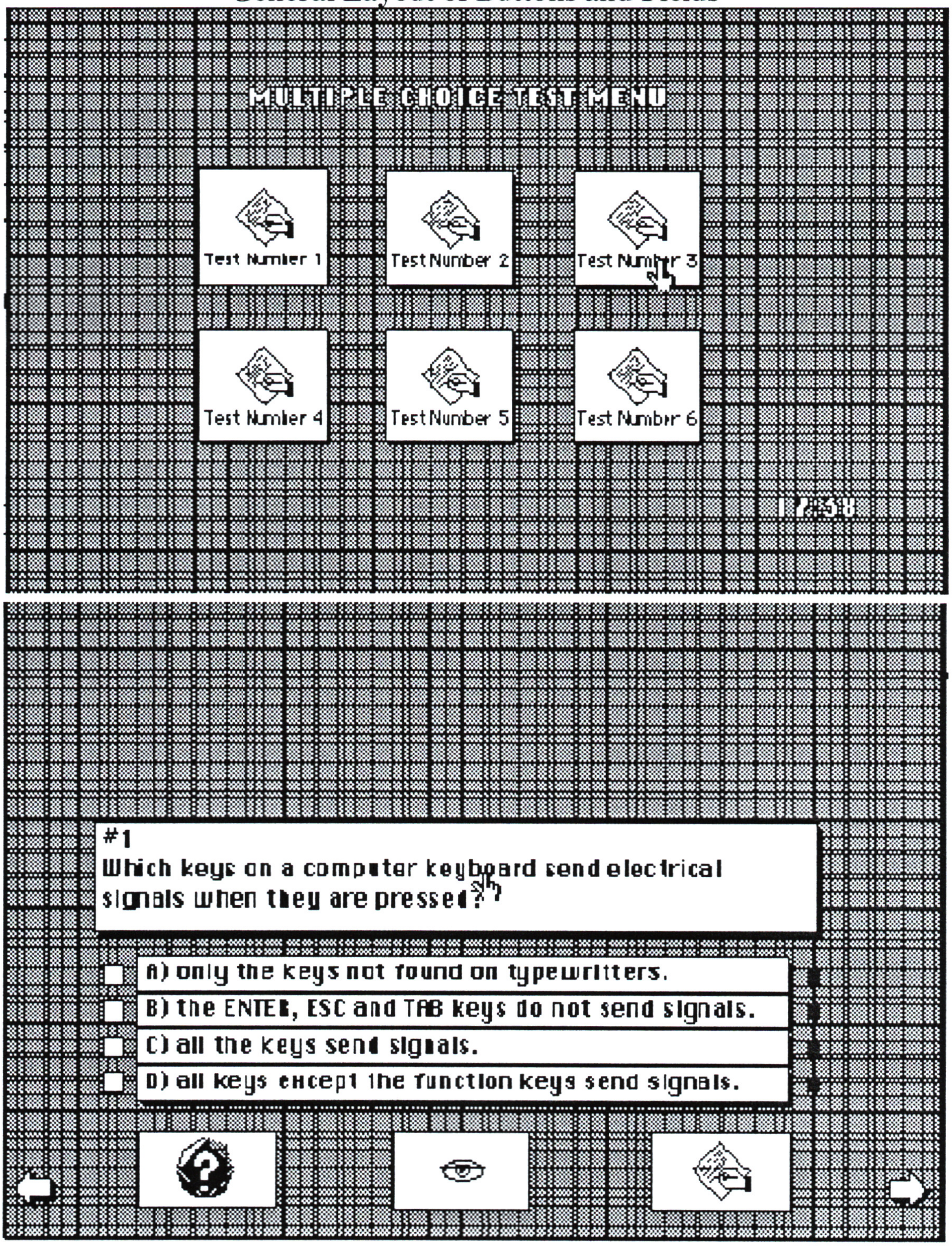
List of Appendices

Appendix A	Computer Screen Prints of the User Interface
Appendix B	Printed Scripts for Software Design Features
Appendix C	Summary of Test Scores and Test Administration Times
Appendix D	Multiple Choice Test Items Used in the Main Study
Appendix E	Pre-Survey Questionaire used in the Main Study
Appendix F	Response Matrix for the Pre-survey Questionaire
Appendix G	Post-Survey Questionnaire used in the Main Study
Appendix H	Response Matrix for the Post-survey Questionaire

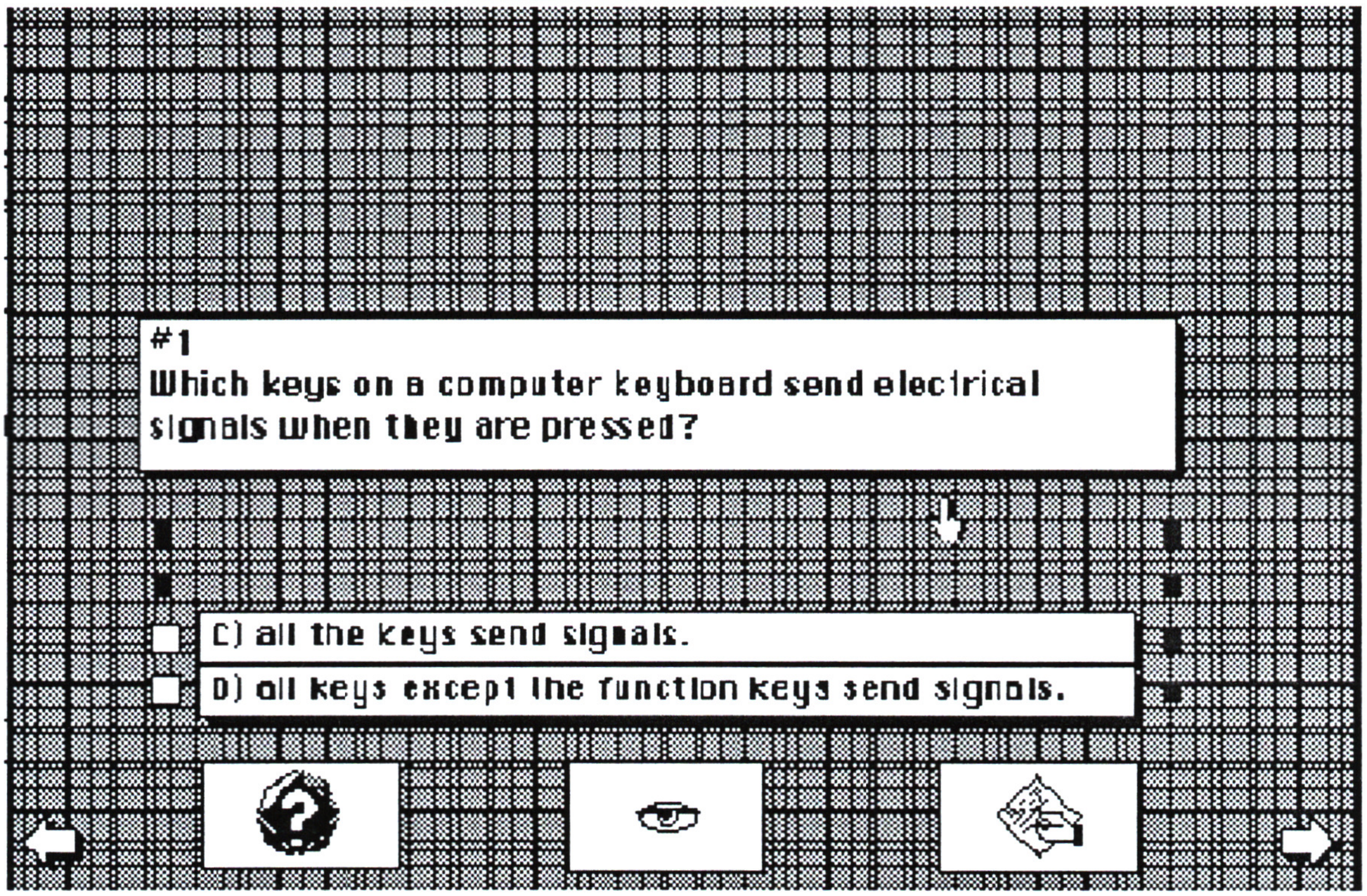
Appendix A

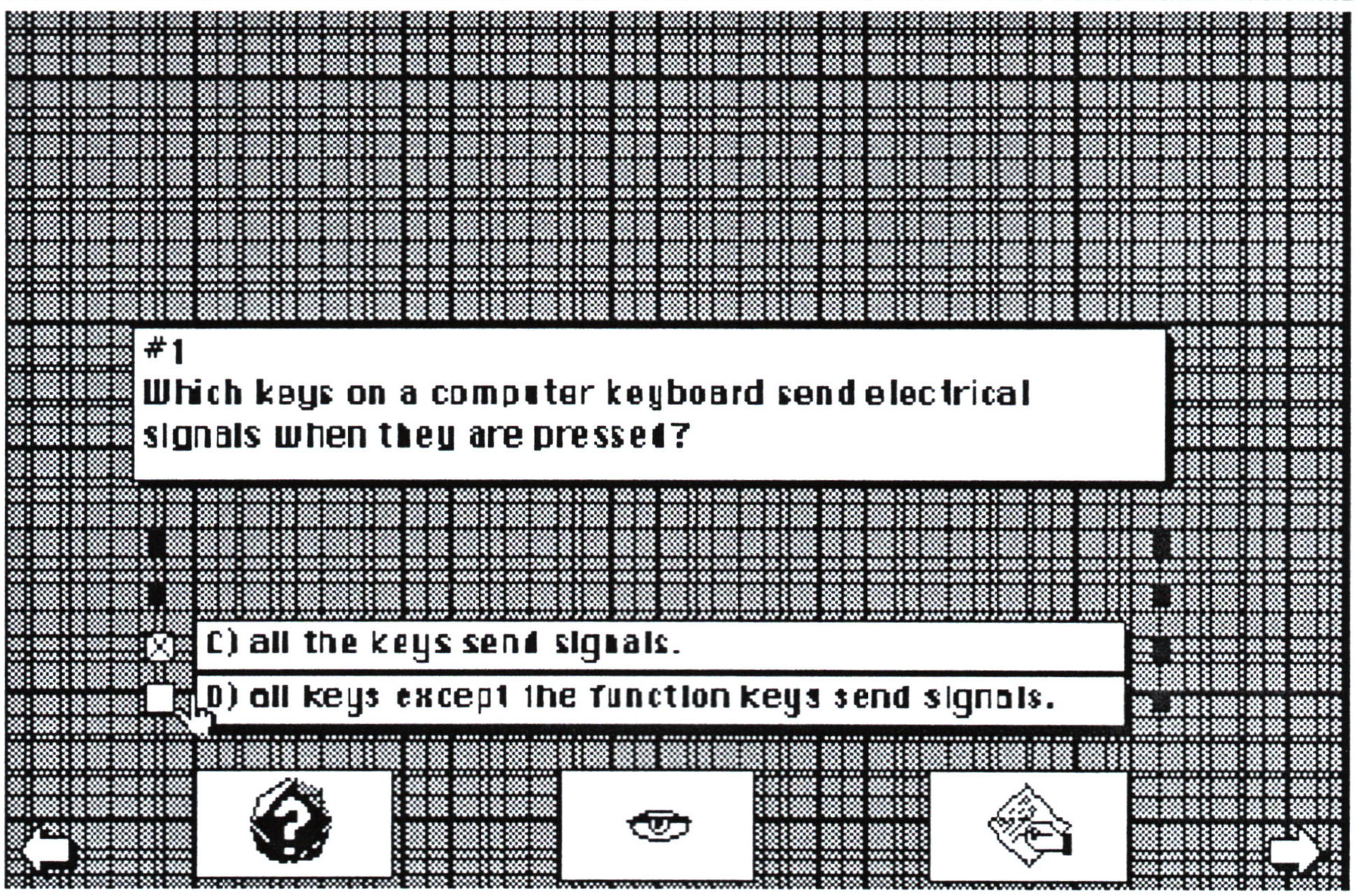
Computer Interface Screen Prints

General Layout of Buttons and Fields

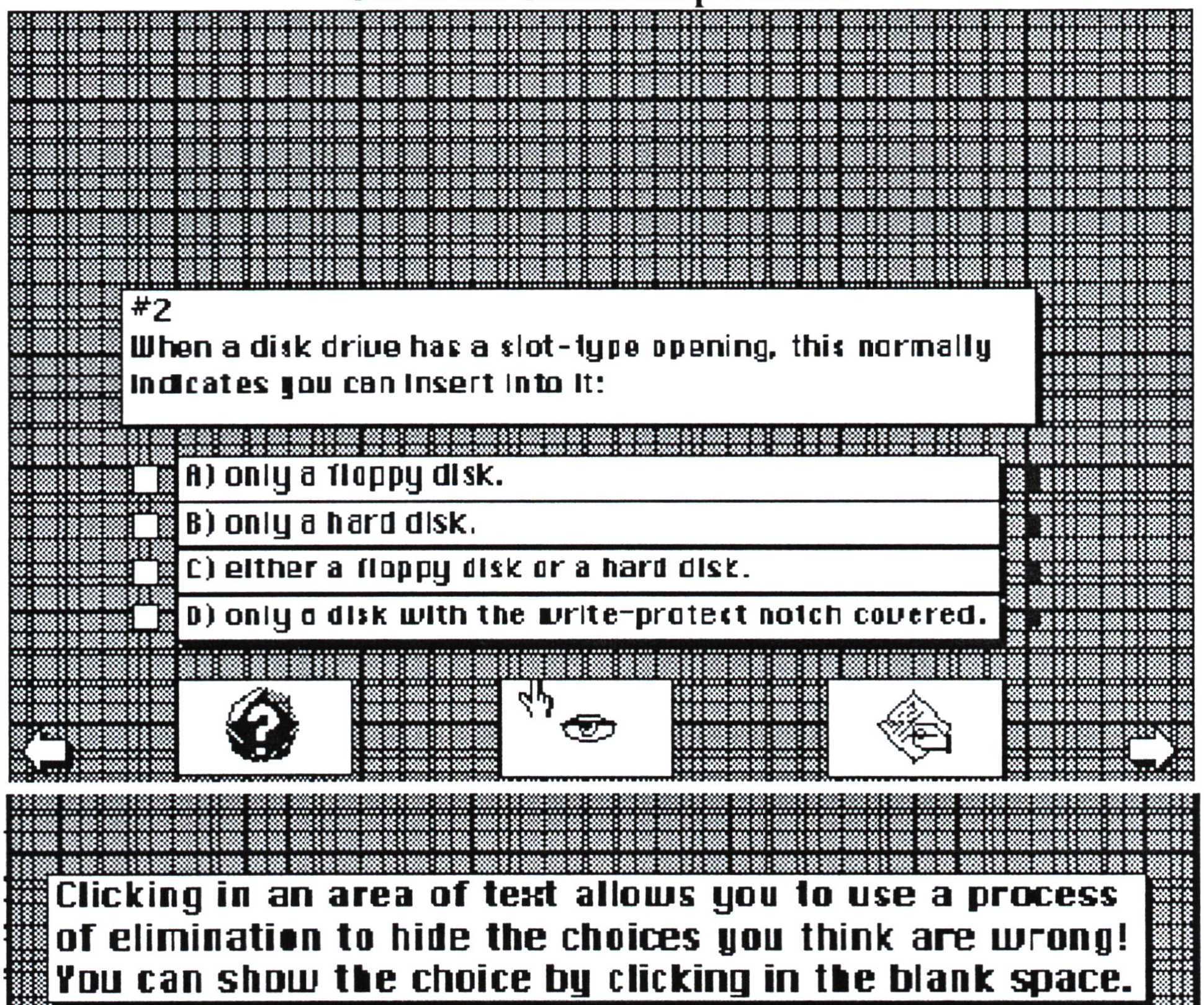


Use of the Elimination and Answer Features

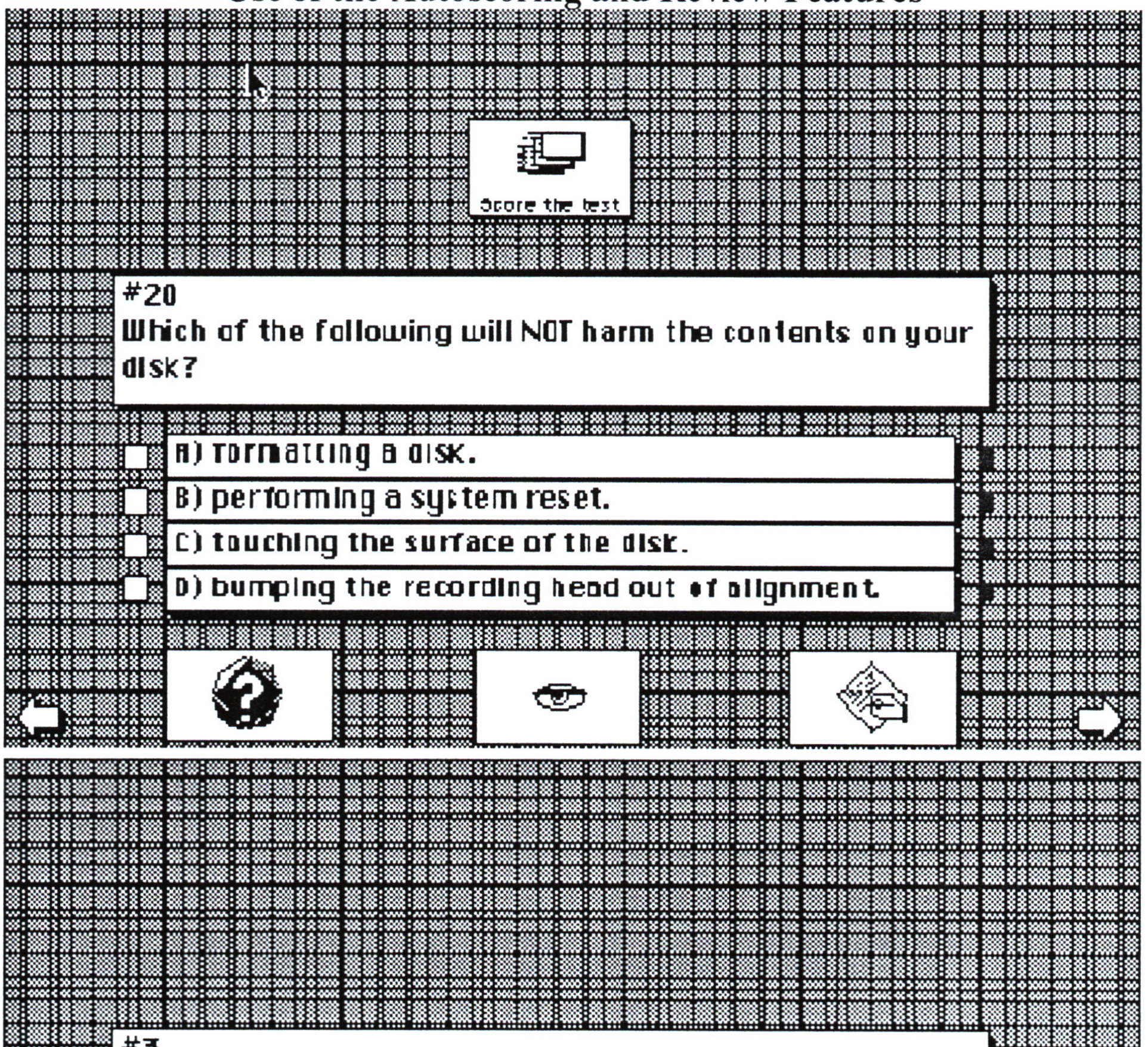




Use of the Online Help Feature



Use of the Autoscoring and Review Features



Appendix B

Printed Scripts for Hypercard Objects

Scripts at the stack level

```
on openStack
  answer "This will take a moment to set up!"
  global name --define the variables for marking the test
  global mark
  etc....
  hide menubar--reset the test for new student
set userlevel to 1
repeat with i = 1 to 20
  set hilite of card button 1 of card i to false
  set hilite of card button 2 of card i to false
  set hilite of card button 3 of card i to false
  set hilite of card button 4 of card i to false
end repeat
--get the student's name
put empty into it
ask "To begin, Please type in your name below ..."
if it is empty then
  ask "You must type in your name!"
end if
put it into name
answer "Are you ready to begin?  " & name with "Yes" or "No"
if it is "No" then --exit the test to the test menu
    go back
  end if
  if it is "Yes" then -- show the first question
    hide card field 1
  end if
end openStack

on closeStack
  hide bkgnd button 6
  show card field 1 of card 1-- block view of questions
  hide card field 1 of card 20
end closeStack

on idle --show the marking button if it is the last question
  get the number of this card
  if it is 20 then
    show bkgnd button 6
  end if
  if it is not 20 then
    hide bkgnd button 6
  end if
end idle
```

Scripts at the card level

```
on openCard
  show bkgnd field 2 --reset the question for a review
  show bkgnd field 3
  show bkgnd field 4
  show bkgnd field 5
  hide bkgnd field 6
  hide bkgnd field 7
  hide bkgnd field 8
  hide bkgnd field 9
  show card button 1
  show card button 2
  show card button 3
  show card button 4
end openCard

on closeCard
  global cardscore1 --score the question right or wrong
  get the hilite of card button 3
  if it is true then
    put it into cardscore1
  else put false into cardscore1
end closeCard
```

Scripts at the button and field level

Script for Answer Button

```
on mouseUp
  if hilite of button 2 is true
  then set hilite of button 2 to false
  if hilite of button 3 is true
  then set hilite of button 3 to false
  if hilite of button 4 is true
  then set hilite of button 4 to false
  set hilite of me to true
end mouseUp
```

Script for Help Button

```
on mouseUp
  play harpsichord a
  show bkgnd field 10
  wait 5 secs
  set hilite of bkgnd button 1 to true
  play harpsichord a
  send mouseUp to bkgnd button 1
  set hilite of bkgnd button 1 to false
  wait 3 secs
  play harpsichord a
  set hilite of bkgnd button 2 to true
  send mouseUp to bkgnd button 2
  set hilite of bkgnd button 2 to false
  hide bkgnd field 10
end mouseUp
```

Script for Arrow Button

```
on mouseUp
  visual effect dissolve
  go next
end mouseUp
```

Script for choice field

```
on mouseUp
  hide me
  hide button 1
  show bkgnd field 6
end mouseUp
```

Appendix C

Summary of Test Scores and Test Administration Times

Test Scores for Tests One and Two

Test One		Test Two	
Computer Group A	Written Group B	Computer Group B	Written Group A
13	15	14	9
12	16	14	12
13	15	20	9
15	16	16	13
12	11	15	13
15	9	10	15
15	14	14	15
12	14	12	10
9	14	17	16
12	10	13	15
18	10	11	18
15	10	10	14
15	15	12	16
8	16	17	14
16	11	10	14
15	15	13	18
14	6	11	17
18	9	9	18
18	19	17	19
17	19	18	20
15	17	18	16
9	17	20	10
15	11	16	19
14	19	20	18
*	17	16	*
sample size			
N=24	N=25	N=25	N=24
mean			
13.96	13.80	14.52	14.92

Test Scores for Tests Three through Six

Test Three		Test Four		Test Five		Test Six	
Computer Group A	Written Group B	Computer Group B	Written Group A	Computer Group A	Written Group B	Computer Group B	Written Group A
14	15	14	14	15	13	14	16
14	16	12	12	17	15	15	16
14	13	10	16	15	13	13	18
15	15	11	9	15	15	15	15
16	11	7	11	13	9	*	13
11	12	9	13	13	10	14	11
17	14	12	13	18	13	15	19
17	18	14	13	17	16	*	17
16	16	10	13	18	11	14	19
12	12	12	9	14	18	17	12
18	15	12	14	18	12	12	14
18	11	7	13	16	13	10	20
16	15	15	13	18	17	19	17
17	13	13	14	10	15	17	15
17	17	14	14	17	16	19	18
18	20	14	16	20	16	17	18
16	17	10	11	13	12	14	14
16	18	18	14	13	19	17	15
16	18	13	13	16	15	17	18
sample size							
N=19	N=19	N=19	N=19	N=19	N=19	N=17	N=19
mean scores							
15.68	15.05	11.95	12.89	15.58	14.11	15.24	16.05

Test Administration Times (in seconds) for Tests Three through Six

Test Three		Test Four		Test Five		Test Six	
Computer Group A	Written Group B	Computer Group B	Written Group A	Computer Group A	Written Group B	Computer Group B	Written Group A
822	1080	2050	745	975	320	1064	620
973	798	1158	1043	751	596	1126	746
1093	810	1131	997	523	1030	793	735
704	810	1470	522	773	1486	657	722
1403	676	1680	599	1235	1265	*	858
985	1705	2297	492	740	1174	1400	536
912	810	776	625	775	968	1166	558
766	780	918	531	712	816	*	934
1461	1278	806	599	717	899	1019	564
814	1406	1254	870	521	1328	707	555
756	405	1544	1082	929	816	1382	912
834	557	1719	722	808	955	1306	923
737	360	819	730	975	870	493	484
1379	900	1220	1070	1205	710	530	890
1137	1080	1118	1295	763	690	620	1365
728	720	769	1080	1551	700	994	860
609	900	715	1270	1152	465	1153	1515
1685	1200	801	1070	588	885	738	1108
1298	540	1145	2065	678	1470	600	945
sample size							
N=19	N=19	N=19	N=19	N=19	N=19	N=17	N=19
mean scores							
1005.05	885.00	1231.05	916.16	861.63	918.05	926.35	833.16

Appendix D

Multiple Choice Test Item Sets

Computer Test Number One

#1
Which keys on a computer keyboard send electrical signals when they are pressed?
A) only the keys not found on typewritters.
B) the ENTER, ESC and TAB keys do not send signals.
C) all the keys send signals.
D) all keys except the function keys send signals.

#2
When a disk drive has a slot-type opening, this normally indicates you can insert into it:
A) only a floppy disk.
B) orrly a hard disk.
C) either a floppy disk or a hard disk.
D) only a disk with the write-protect notch covered.

#3
Dot matrix printers show letters on paper by:
A) magnetizing spots on the paper which attract ink
B) rushing small wires or pins against an ink ribbon.
C) releasing ink onto the paper through matrix ports.
D) using a tractor head which rolls across the paper.

#4
What is the name of the plug-in type openings that allow the processing unit to be connected to other parts of the system?
A) heads.
B) codes.
C) devices.
D) ports

#5
Which part of the computer retrieves instructions one at a time and executes them?
A) the port.
B) the central processing unit.
C) the memory.
D) the monitor.

#6
System software is one of the four main types of software. Which of the following is NOT a main type?
A) utility software.
B) application software.
C) tutorial software.
D) programming languages.

#7
What is the difference between large and small computers?
A) large computers don't use input/output devices.
B) L. computers do many instructions simultaneously.
C) small computers are slower and have less memory.
D) only small computers need storage devices.

#8
Which of the following is an example of computer software?
A) a site license.
B) a floppy disk.
C) a tutorial.
D) a printed letter.

#9
When data is typed in on a keyboard, how is it sent to the processing unit?
A) as electrical signals.
B) the central processing unit.
C) in a dot matrix form.
D) as a set of magnetized spots.

#10
When we speak of a microcomputer's system, we mean:
A) the computer's programs.
B) the computer's combined hardware and software.
C) the way the computer processes data.
D) the computer's devices for storage and display.

#11
To say that a program is loaded means that:
A) it is full of instuctions and can do many tasks.
B) its instructions are executed one by one.
C) it is copied from the disk into memory.
D) it shows up on the screen of the monitor.

#12
Which of the following allows you to create your own computer instructions (software)?
A) a word processing program.
B) a programming language.
C) a site licence.
D) a tutorial.

#13
In order to make certain that the files on your disk will never be changed you should:
A) cover the write protect notch.
B) always use a hard disk.
C) use the cleaning solvent at least once a month.
D) never put a disk in the drive while the light is on.

#14
The purpose of the tractor feed mechanism on a printer is to :
A) push the printing head across the paper.
B) allow you to move the paper by hand.
C) advance paper one line at a time.
D) pull continuous paper through the printer.

#15
The term firmware refers to software that is :
A) stored on a hard disk.
B) permanently installed in memory circuits.
C) part of the system software.
D) well docu me nted.

#16
One example of documentation is:
A) a programming language.
B) a software manual.
C) a piece of firmware.
D) a copyright.

#17
Through which of the following is new data entered into the computer?
A) the disk drive.
B) the monitor.
C) the keyboard.
D) the memory.

#18
A surge suppressor is used to:
A) avoid amage from strong electrical currents.
B) keep paper from moving too fast in the printer.
C) control the movement of the disk drive head.
D) stop you from pressing too hard on the keyboard.

#19
Occasionally a disk drive is out of alignment. This means that:
A) one drive is not above the other.
B) the drive was turned sideways in the casing.
C) the recording head cannot put data on a disk.
D) the disk will not fit into the drive.

#20
Which of the following will NOT harm the contents on your disk?
A) formatting a disk.
B) performing a system reset.
C) touching the surface of the disk.
D) bumping the recording head out of alignment.

Computer Test Two

#1
In reference to disk storage, what are tracks?
A) the list of all files that are stored on the disk.
B) wedge shaped areas which contain the data.
C) circular lines of magnetized spots.
D) a place on disk where file specifications are kept.

#2
The disk directory contains at least:
A) the names of the files stored on disk.
B) the locations in memory where files are stored.
C) the list of defaults and wildcards for the system.
D) the commands you may use to access the disk.

#3
Which of the following is NOT used by the system to identify a specific file on disk?
A) the extension.
B) the name of the file
C) the name of the drive.
D) the label.

#4
When you are naming a disk file, what rules must you follow?
A) different rules for different disk sizes.
B) those specific to the system you are using.
C) the rules given in the default settings.
D) only letters and no more than 8 characters.

#5
When you are typing in an MS/DOS file specification,
which of the following could be used as a file name?
A) DATA-FILE
B) CASYAT.BAT
C) EXE MYFILE
D) TXT.YOURFILE

#6
In MS/DOS the wildcard charcters are:
A) * and ?
B) A: and B:
C) letters and digits
D) A> and B>

#7
The process of loading the operating system into the computer's memory is called:
A) copying
B) formatting
C) programming
D) booting

#8
When a program a group of data items is stored on a disk it is called?
A) a specification
B) a sub-directory
C) a block
D) a file

#9
When the cursor is blinking at the "A>" prompt this means:
A) the system needs to be loaded into memory
B) the computer is waiting for a command
C) the disk has no file in its directory
D) the disk needs to be fom7atted

#10
What happens when you format a disk?
A) The files are arranged irrto alphabetical order.
B) The disk is divided into sectors and tracks.
C) The system is loaded from disk into memory.
D) The computer shows you what files are on disk.

#11
As soon as an MS/DOS computer is switched on:
A) it shows a svstem prompt on the screen .
B) it loads the disk directory into memory .
C) it looks for the operating system on a disk .
D) it formats the disk in drive A.

#12
When using MS/DOS, what is the correct command to type when the prompt is "A>", the system disk is in drive A, and the disk to be formatted is in drive B?
A) FORMAT B:
B) FORMAT DISK B
C) FORMAT A>B
D) FORMAT A

#13
The prompt "A>" is on the screen and the disk containing a file called EXAM.TXT is in the B drive. What do you type to duplicate that file onto the disk in the A drive?
A) COPY B:EXAM.TXT A:
B) MOVE EXAM.TXT TO A
C) COPY EXAM.TXT B>A
D) MOVE EXAM.TXT FROM B: TO A:

#14
The prompt is "B>" and you want to erase all the files on the disk in the A drive whose names start with the letters ABC. What command do you type?
A) DEL A:ABC*
B) DEL ALL A:ABC*
C) DEL A:ABC.*
D) DEL A:ABC*.*

#15
The phrase "system reset" means:
A) putting the computer hardware back in place
B) assembling the computer system
C) re-booting the system
D) copying the operating system back onto the disk

#16
When the system prompt "A>" is on the screen, what do you type to show the names of all the files in drive B?
A) DIR B:
B) DIR
C) DIR A>B
D) DIR B.FILES

#1 7 When you have the "A-" prompt on the screen, what do you type to execute the file called ADDONE which is on the disk in A drive and has the extension EXE?
A) EXECUTE ADDONE
B) EXE A:ADDONE
C) RUN ADDONE
D) ADDONE.EXE

#18
Which of the following best describes the command MOVE ?
A) It erases a file from 1 disk copying it to another.
B) It is used to copy file from one disk to another.
C) It is not a valid MS/DOS command.
D) It loads a file from the disk to the memory.

#19
If the prompt "A>" is on the screen and a blank formatted disk is in the B drive, what existing files would be affected by the command COPY MINE.* B: ?
A) every file starting with the letters MINE
B) only the one file called MINE.*
C) whatever files on the B drive named MINE
D) all files on the A drive with the file name MINE

#20
A file has the extension COM and you want to change its file name from OLDNAME to NEWNAME. You should type:
A) REN NEWNAME.*
B) REN OLDNAME.* NEWNAME.*
C) REN OLDNAME.COM NEWNAME.COM
D) REN NEWNAME.COM OLDNAME.COM

Computer Test Number Three

#1
Why is the microcomputer also callea a personal computer?
A) It started with the IBM PC.
B) It is user friendly and easy to work with.
C) It can be used by only one person at a time.
D) It is useful for an individual but not a business.

#2
Which of the following is TRUE for every type of keyboard ?
A) It is part of the processing unit.
B) It connects directly to the monitor.
C) It must be used to get irrstructions into memory.
D) It sends electrical signals to the computer.

#3
Why can data be stored on floppy disks and hard disks?
A) The recording head has pins that form images.
B) Tiny points have static electricity put on them.
C) They are coated with a magnetic substarrce.
D) The computer can scratch marks on their surfaces.

#4 What button must you press in order to stop the printer from receiving more signals from the computer?
A) the off-line button.
B) the form feed button.
C) the tractor feed button.
D) the power-on button.

#5 What part of the computer is used to store the instructions and data on which the computer is currently working?
A) the memory.
B) the software.
C) the diskdrive.

#6
In reference to disk storage, what are tracks?
A) the list of all files that are stored on the disk
B) wedge shaped areas which contain the data
C) circular lines of magnetized spots
D) a place on disk where file specifications are kept

#7
In MS/DOS the wildcard characters are:
A) * and ?
B) A: and B:
C) letters and digits
D) A> and B>

#8
The process of of loading the operating system into the computer's memory is called:
A) copying
B) formatting
C) programming
D) booting

#9
When the cursor is blinking at the system prompt "A-", this means:
A) the system must be loaded into memory
B) the computer is waiting for a command
C) the disk has no files in its directory
D) the disk needs to be formatted

#10
What happens when you format a disk?
A) the files are arranged into alphabetical order.
B) the disk is divided into sectors and tracks.
C) the system is loaded form disk into memory.
D) the computer lists the files that are on the disk.

#11
Which keys on a computer keyboard are programmable?
A) Only the cursor keys.
B) Only the function keys.
C) All the keys.
D) All the keys except the letters and digits.

#12
Which key normally forces all the letters of the alphabet to become uppercase letters?
A) PGUP.
B) CAPSLOCK.
C) ALT.
D) F1.

#13
Which key normally erases the character to the left of the cursor?
A) LEFT ARROW.
B) BACKSPACE.
C) HOME.
D) F10.

#14
Which key normally pressed to signal the end of a paragraph?
A) PGDN.
B) END.
C) BACKSPACE.
D) ENTER.

#15
What key do you press when you wish to use the number pad for entering digits rather than for controlling the cursor?
A) NUMLCK.
B) SHIFT.
C) ESC.
D) BREAK.

#16
What will happen when you press the letter "M" on a computer keyboard?
A) The disk drive will store the letter until needed.
B) Different actions each for hard and floppy disks.
C) Whatever action is programmed for the computer.
D) The letter "M" will be displayed on the monitor.

#17
Which of the following keys normally does NOT move the cursor?
A) DEL.
B) END.
C) PGUP.
D) HOME.

#18
What is the use of the Fl function key?
A) It is used to halt the execution of a program.
B) It moves the cursor to the top of the screen.
C) It is used with the letters to give more options.
D) It has no standard use.

#19
Which keys are used in combination with other keys to double the number of codes you can enter?
A) CTRL and ALT.
B) INS and DEL.
C) PGUP and PGDN.
D) ESC and ENTER.

#20
Which key is often used to halt an action being performed by the program?
A) END.
B) ESC.
C) DEL.
D) BACKSPACE.

Computer Test Number Four

#1
Which of the following is NOT software:
A) programs.
B) documentation.
C) instructions.
D) firmware.

#2
During the processing cycle, which one of the following actions produces information?
A) input
B) processing
C) storage
D) output

#3
Which of the following is an example of computer hardware?
A) a programming language.
B) a port.
C) a copy right.
D) a back-up.

#4
The computer's processing unit has two parts - the CPU and:
A) the memory.
B) the disk drive.
C) the operating system.
D) the program.

#5
When using MS/DOS, what is the correct command to type when the prompt is "A>", the operating system is in drive A and the disk to be formatted is in drive B?
A) FORMATB:
B) FORMAT DISK B
C) FORMAT B>A
D) FORMAT A

#6
The prompt "A>" is on the screen and the disk containing a file EXAM.TXT is in the B drive. What do you type to make a duplicate of the file on the disk in A drive?
A) COPY B:EXAM.TXT A:
B) MOVE EXAM.TXT TO A
C) COPY EXAM.TXT B>A
D) MOVE EXAM.TXr FROM B: TO A:

#7
The prompt is "B>" and you want to erase all the files in the A drive whose names start with the letters ABC. What command would you type?
A) DEL A: ABC*
B) DEL ALL A:ABC*
C) DEL A: ABC.*
D) DEL A: ABC*.*

#8
The word "tool" is often used in the Works tutorials.
When the word "tool" is used in this fashion it refers to:
A) special keys for insertlng, deleting or editing text.
B) any of the commands found in the Works menu.
C) one complete applicatlon within Works.
D) an action you can perform on a paragraph.

#9
How do you access a command that is in a Works menu?
A) press a letter matching the 1 st letter in the name.
B) move the cursor to the menubar using ARROWKEYS.
C) highlight the command using TAB then ENTER.
D) press ALT then one of the highlighted letters.

#10
What is the purpose of the status line in Works?
A) It warns you when you have made an error.
B) It tells you what flie you are using.
C) It shows you all the commands you can select.
D) It prompts you on the purpos(e of each command.

#11
What is the difference between selecting a command and selecting some text?
A) Selecting a command executes that command.
B) Selecting text is done with the control key.
C) Selecting a command means to find it on the disk.
D) When you select some text your file is changed.

#12
When would you use the "Save As" command instead of the "Save" command?
A) You want a 2nd copy of the file with a new name.
B) When you want to put the file on a hard disk.
C) When it is a new file and you need to name it.
D) When the file name is valid but lacks an extension.

#13
When would you use the "Open" command instead of the "New" command?
A) When you want to type a letter.
B) When you want to remove the disk from the drive.
C) Whenever you want to work with more than 1 file.
D) When the file you want already exists on the disk.

#14
Which key is normally prelsed to move or scroll down one row in your spreadsheet file?
A) PGDN.
B) END.
C) BACKSPACE.
D) DOWN ARROW.

#15 What key do you press when you wish to use the number pad for entering digits rather than for controlling the cursor?
A) NUMLCK.
B) SHIFT.
C) ESC.
D) BREAK.

#16
What will happen when you move to a cell in a spreadsheet file then press enter?
A) The disk drive will store the new entry.
B) The previous entry will replaced by a blank cell.
C) Nothing
D) The return character will be displayed in the cell.

#17
Which of the following keys will NOT move the cursor more than one row at a time in a spreadsheet file?
A) DOWN ARROW.
B) END.
C) PGUP.
D) HOME.

#18
What is the use of the F2 function key in the Works spreadsheet?
A) It is used to halt the execution of a program.
B) It moves the cursor to the top of the screen.
C) It is used for editing an entry within a cell.
D) It has no standard use.

#19
Which keys are used to cancel or execute specific commands in a Works spreadsheet ?
A) CTRL and ALT.
B) INS. and DEL.
C) PGUP and PGDN
D) ESC and ENTER.

#20
Which key is used to bring the user to the top of a file?
A) HOME.
B) ESC.
C) DEL.
D) BACKSPACE.

Computer Test Number Five

#1
Without turning the printer off, what do you do to advance the paper an inch?
A) Press the form-feed button.
B) Use the line-feed button.
C) Use your hand to roll the paper forward one inch.
D) Push the feed-ty pe lever to pressure feed.

#2
The normal device that the computer uses to ask the human user for instructions is:
A) the monitor.
B) the disk drive.
C) the processing unit.
D) the printer.

#3
Why does the microcomputer need a disk drive?
A) It executes the program that is stored on disk.
B) Data only in memory is lost when the power is off.
C) Data entered at the keyboard must be stored.
D) The CPU obtains instructions from the diskdrive.

#4
What is the major difference between the floppy disk and the hard disk?
A) Data on the hard disk is permanent.
B) Floppy disks are coated with a magnetic film.
C) A hard disk is permanently enclosed in the drive.
D) The hard disk is always inside the computer.

#5
Which of the following does NOT involve electrical signals?
A) Insructions are present on the disk.
B) Characters are sent to the printer.
C) A program is copied into memory.
D) Data is entered on the keyboard.

#6
When a program' s instruction are being carried out by the processing unit, you say the program is being:
A) processed.
B) loaded.
C) backed up.
D) executed.

#7
Which of the following refers to the basic programs needed by the computer for such things as keeping track of what is in memory and getting keyboard signals?
A) programming languages.
B) utility software.
C) system software.
D) application software.

#8
The disk is organized into pie-shaped or wedge shaped areas called:
A) sectors.
B) files.
C) tracks.
D) sub-directories.

#9
In order to keep track of where files are located on the disk the system uses a list of files called the:
A) directory.
B) file specification.
C) label.
0) extension.

#10
Which is a valid MS/DOS drive name that you could include in a file specification?
A) B:
B) A>
C) #1
D) D2

#11
When you give a file specification to the computer, what is meant by a default?
A) the name of the drive.
B) an asterisk ,"*" or a question ma rk" ? ".
C) what the computer will use if you don't specify.
D) an error message that is displayed.

#12
Which of the following shows the correct order for the parts of a full file specification?
A) filename: drive.extension
B) filename.extension:drive
C) drive:filename. extension
D) drive:extension. filename

#13
Another name for the operating system is:
A) the transient commands.
B) the resident program.
C) the system software.
D) the command language.

#14
The phrase "system reset" means:
A) putting the computer hardware back in place.
B) assembling the computer system.
C) re-booting the system.
D) copying the operating system back onto the disk.

#15
When would you use the "Save As" command?
A) to create a new file.
B) to save an extra copy of a file under a new name.
C) to save a copy of a file on another disk.
D) to rename a file.

#16
Spreadsheet files may contain the following:
A) Text, paragraphs and numbers.
B) Fields, Records and Cells.
C) Formulas, Text or Numbers.
D) Records, Columns and Cells.

#17
Which of the following keys will NOT move the cursor more than one row at a time in a spreadsheet or database file?
A) DOWN ARROW.
B) END.
C) PGUP.
D) HOME.

#18
What is the use of the ESC key in the Works spreadsheet?
A) It is used to halt the execution of a command.
B) It moves the cursor to the top of the screen.
C) It is used for editing an entry within a cell.
D) It has no standard use.

#19
Which key is used to execute specific commands in Works?
A) ALT.
B) DEL.
C) PGDN
D) ENTER.

#20
Which key is used to bring the user to the end of a file?
A) HOME.
B) END.
C) DEL.
D) BACKSPACE.

Computer Test Number Six

#1
Which of the following stores data and instructions outside of memory?
A) the CPU.
B) the disk drive.
C) the processing unit.
D) the monitor.

#2
How much data will normally be erased when the computer's electrical power is cut off?
A) both what is in memory and what is on any disk.
B) what is in memory and what is on a floppy disk.
C) only what is in memory.
D) only what is on disk.

#3
When putting the 5 & 1 /4 inch disk into the drive, what is the proper position for the disk?
A) The side with the label on it should be up.
B) The side with the label on it should be down.
C) the write protect notch should go in first.
D) the write protect notch should go in last.

#4
When you are printing a long document on continuous feed paper, which of the following should you NOT do?
A) Select the on-line mode.
B) Leave the printer unattended.
C) Have the power switched on.
D) Switch from pressure feed to tractor feed.

#5
If a program has a copywrite, this means:
A) you must cover the write protect notch.
B) you cannot use it on many computers at one time.
C) the program is impossible to copy.
D) it is illegal to copy it without permission.

#6
Which of the following is NOT important in keeping track of your data?
A) Put each data file on a separate disk.
B) Attach a label to each disk.
C) Erase a data file when it is no longer useful.
D) Organize your disk's contents into subgroups.

#7
What is the name of the plug-in type openings that allow the processing unit to be connected to other parts of the system?
A) heads.
B) codes.
C) devices.
D) ports.

#8
A surge suppressor is used to:
A) avoid damage from strong electrical currents
B) keep the paper from moving ahead too fast.
C) control the disk drive's recording head.
D) stop you from presslng too hard on the keyboard.

#9
What is the next thing the computer does after the operating system is loaded into memory?
A) It performs a system reset.
B) It shows a system prompt on the screen.
C) It lookI for a file on the disk.
D) It initializes the disk.

#10
A disk's directory contains:
A) the names of aII the files on it
B) the operating system instructions.
C) all the MS/DOS commands.
D) all the data you have entered and stored on disk

#11
When the system prompt "A>" is on the screen, what would you type to show the names of all the files in drive B?
A) DIR B:
B) DIR
C) DIR A>B
D) DIR B.FILES

#12
When you have the "A>" prompt on the screen, v\r'hat command wou\d you type to execute the file called ADDONE.EXE which is on the disk in the A drive?
A) EXECUTE ADDONE
B) EXE A:ADDONE
C) RUN ADDONE
D) ADDONE.EXE

#13
The process by which the surface of the disk is prepared for use by dividing it into wedge-shaped sections and concentric circles is called:
A) booting the disk.
B) formatting the disk.
C) defaulting the disk.
D) organizing the disk.

#14
Which of the following best describes the MS/DOS command MOVE ?
A) it erases a file from a disk, copying it to another.
B) it copies a file from one disk to another.
C) it is not a valid MS/DOS command.
D) it loads a file from the disk into the memory.

#15
If the prompt "A>" IS On the screen and a blank formatted disk is in the B drive, what existing files would be affected by the command COPY MINE.* B:
A) all files whose name starts with the letters MINE
B) only the one file called MINE.*
C) all the files on the B drive named MINE
D) all the files on the A drive named MINE

#16
A file has the extension COM and you want to change its filename from OLDNAME to NEWNAME. You should type:
A) REN NEWNAME.*
B) REN OLDNAME.* NEWNAME.*
C) REN OLDNAME.COM NEWNAME.COM
D) REN NEWNAME.COM OLDNAME.COM

#17
Which of these characters does WORKS use to recognize the end of a paragraph in a word processing file?
A) RETURN
B) PERIOD
C) TAB
D) SPACE

#18
What is the use of the ESC key in the Works spreadsheet?
A) It is used to halt the execution of a command.
B) It moves the cursor to the top of the screen.
C) It is used for editing an entry within a cell.
D) It has no standard use.

#19
When in the FORM view of a data base file in WORKS, which key allows you to move between field and records?
A) ALT.
B) DEL.
C) TAB.
D) ENTER.

#20
When moving text or numbers betwen the different types of tools found in WORKS, this is called:
A) Integrating.
B) Wordprocessing.
C) Using the Spreadsheet.
D) Using the Database.

Appendix E

Pre-Survey Questionnaire

Please take a few minutes to answer the following 10 questions as honestly as you can. You do not need to sign your name.

1. Are you male or female? M F

2. What is your age? _____

3. How many years have you been out of highschool? _____

4. Briefly, what types of work experience do you have?

_____

_____

5. Do you feel you do well on multiple choice tests? Y N

6. How much experience with computers do you have? _____

7. Have you taken a test on a computer before? Y N

8. In your opinion will you prefer the written or computer test? _____

9. Do you believe that computer tests will improve your grade? _____

10. What are the advantages and disadvantages of computerized tests?

_____

_____

_____

_____

Thankyou for taking the time to fill out this survey.

Appendix F

Response Matrix for the Pre-Survey Questionnaire

Question No.

#1	#2	#3	#4	#5	#6	#7	#8	#9	#10 a	#10 b
F	19	0.5	service	Y	0	N	n/a	Y	faster	confusing
F	19	1	labour	Y	0	N	W	Y	neater	confusing
F	19	1	sports	Y	2	N	C	Y	elimination feature	eye fatigue
F	20	2	service	Y	5	N	W	N	instant scoring	impersonal
F	20	2	service	N	0	N	W	N	neater	inconvenient
F	20	2	labour	Y	0	N	n/a	n/a	convenient	inexperience
F	20	2	service	N	0.5	N	C	n/a	elimination feature	inexperience
F	20	4	service	N	0	N	W	Y	increased feedback	inexperience
F	21	4	service	Y	2	N	n/a	n/a	learning	inexperience
F	21	4	service	Y	3	N	C	Y	clearer/easier	intimidating
F	22	4	service	Y	0	N	W	Y	convenient	intimidating
F	22	4	office	n/a	0	N	n/a	n/a	easier	intimidating
F	22	4	service	Y	0	N	C	n/a	faster	intimidating
F	22	5	service	Y	0.5	N	W	Y	learning	intimidating
F	22	5	service	Y	0	N	C	n/a	neater	less review
F	23	6	industry	Y	5	Y	C	N	easier	n/a
F	23	6	labour	Y	0	N	C	n/a	easier	n/a
F	24	6	service	N	2	N	C	Y	easier	n/a
F	24	6	service	Y	0	N	C	Y	easier/convenient	n/a
F	24	8	office	Y	4	N	C	Y	easier/convenient	n/a
F	24	8	service	Y	0	N	W	N	faster/easier	n/a
F	27	8	service	Y	0	N	n/a	n/a	faster/easier	n/a
F	27	9	service	Y	0	N	C	Y	faster/easier	n/a
F	28	10	industry	Y	0	N	C	n/a	instant scoring	n/a
F	28	10	service	Y	0	N	C	n/a	instant scoring	n/a
F	28	10	educatio	Y	0.5	N	C	n/a	instant scoring	n/a
F	29	12	service	Y	0	N	C	Y	learning	n/a
F	30	12	service	Y	1	Y	C	Y	learning	n/a
F	30	12	service	N	0	N	C	n/a	less intimidating	n/a
F	30	12	labour	Y	0.5	N	W	N	n/a	n/a
F	31	13	industry	Y	0	N	C	n/a	n/a	n/a
F	31	14	service	Y	1	N	C	n/a	n/a	n/a
F	31	14	industry	n/a	0	N	n/a	n/a	n/a	n/a
M	31	14	industry	Y	0	N	n/a	n/a	n/a	n/a
M	32	14	service	N	0.5	N	n/a	n/a	n/a	n/a
M	32	15	health	Y	0	N	W	n/a	n/a	n/a
M	33	15	service	Y	0	N	W	n/a	n/a	n/a
M	33	15	health	N	0	N	C	Y	n/a	n/a
M	33	17	office	N	0	N	C	Y	n/a	n/a
M	34	17	industry	Y	0	N	C	Y	n/a	n/a
M	35	19	service	Y	0	N	C	Y	n/a	n/a
M	35	19	service	Y	0	N	C	Y	n/a	n/a
M	35	20	labour	Y	5	N	n/a	Y	n/a	n/a
M	36	20	service	Y	0	N	W	Y	n/a	n/a
M	37	20	service	Y	0	N	W	Y	n/a	n/a
M	39	21	industry	Y	0	N	C	n/a	neater	n/a
M	43	23	industry	Y	5	N	C	N	neater/faster	n/a
M	49	25	educatio	Y	0	N	C	Y	neater/faster	n/a
M	57	30	office	n/a	30	N	n/a	Y	neater/faster	n/a
M	n/a	40	service	Y	0	N	C	Y	neater/fun	n/a

Appendix G

Post-Survey Questionnaire

FOLLOW-UP SURVEY

Please take a few minutes to answer the following 10 questions as honestly as you can. You do not need to sign your name.

1. Are you male or female? M F

2. What is your age? _____

3. How many years have you been out of highschool? _____

4. In your opinion, did you prefer the written or computer form of the tests?

_____

5. Do you believe that writing computerized tests improved your grades?

_____

6. Did you find the elimination feature of the computerized test useful?

_____

7. After taking a computerized test, did you review your answers before having the computer mark the test?

_____

8. After the computerized test was marked, did you review your errors?

_____

9. In your opinion, what are the advantages of writing computerized tests?

_____

_____

10. In your opinion, what are the disadvantages of writing computerized tests?

_____

_____

Thankyou for taking the time to fill out this survey.

Appendix H

Response Matrix for Post-Survey Questionnaire

Quest

#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
F	49	30	C	Y	n/a	n/a	n/a	n/a	confusing
F	20	2	C	N	n/a	n/a	n/a	elimination feature	eye fatigue
F	43	24	n/a	N	Y	Y	Y	faster/simpler	feeling rushed
M	19	0.5	C	N	n/a	n/a	n/a	instant scoring	impersonal
F	20	3	C	N	Y	Y	N	appropriate for	n/a
F	32	18	C	N	Y	Y	Y	appropriate for	n/a
M	22	n/a	C	Y	N	Y	N	easier/faster	n/a
F	29	12	C	n/a	n/a	n/a	n/a	easier/faster	n/a
F	57	40	C	Y	n/a	n/a	n/a	easier/faster	n/a
F	23	4	C	Y	Y	Y	Y	easier/fun	n/a
F	32	17	C	N	Y	Y	Y	easier/fun/eliminat	n/a
F	22	4	C	Y	n/a	n/a	n/a	easier/instant	n/a
F	28	9	C	n/a	N	Y	N	easier/simpler	n/a
F	30	15	n/a	n/a	Y	Y	Y	elimination feature	n/a
M	19	n/a	C	Y	Y	Y	Y	elimination feature	n/a
M	27	10	C	Y	Y	Y	Y	elimination/review	n/a
F	22	4	n/a	n/a	Y	Y	N	faster	n/a
M	27	8	C	N	N	Y	Y	faster	n/a
M	20	3	C	N	Y	Y	Y	fun/motivating	n/a
M	28	8	C	Y	Y	Y	Y	fun/motivating	n/a
F	35	19	C	Y	Y	Y	Y	immediate	n/a
F	30	14	C	N	Y	N	Y	instant scoring	n/a
F	36	n/a	C	Y	Y	N	Y	instant scoring	n/a
F	31	15	C	Y	N	Y	Y	instant scoring	n/a
F	33	17	C	Y	Y	Y	Y	instant scoring	n/a
F	40	23	C	Y	Y	Y	Y	instant scoring	n/a
F	31	12	C	n/a	n/a	n/a	n/a	more interesting	n/a
M	24	6	W	N	Y	N	Y	n/a	n/a
F	33	17	W	N	Y	N	Y	n/a	n/a
M	30	14	n/a	N	Y	Y	Y	n/a	n/a
M	35	20	W	N	Y	Y	Y	review feature	n/a
F	20	5	C	N	Y	Y	Y	review/elimination	n/a
M	21	3	C	Y	Y	Y	Y	simpler/instant	n/a
F	28	10	W	n/a	N	Y	Y	instant scoring	no hard copy
F	29	10	C	N	n/a	n/a	n/a	easier/faster	too fast
F	24	8	C	Y	Y	Y	Y	elimination feature	too fast
F	22	5	C	N	Y	Y	Y	clearer	unfamiliar
M	20	1	C	Y	Y	Y	Y	easier	unfamiliar
F	32	14	W	N	Y	Y	Y	faster	unfamiliar
F	n/a	19	C	N	Y	N	Y	instant feedback	unfamiliar
M	24	5	W	N	Y	N	Y	review feature	unfamiliar
F	20	4	W	N	Y	Y	Y	elimination feature	waiting for
F	33	15	W	N	Y	Y	Y	instant scoring	waiting for

VITA

Surname: Zandvliet Given Names: David Bryan
Place of Birth: Toronto, Ont. Date of Birth: October 19, 1962

Educational Institutions Attended:

University of Victoria	1991 to 1993
University of Victoria	1987 to 1988
University of Guelph	1981 to 1987

Degrees Awarded:

B.Sc. (Honours)	University of Guelph	1987

PARTIAL COPYRIGHT LICENSE

I hereby grant the right to lend my thesis to users of the University of Victoria Library, and to make single copies only for such users or in response to a request from the Library of any other university, or similar institution, on its behalf or for one of its users. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by me or a member of the University designated by me. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Title of this Thesis: A Comparison of the Computer Administered and Written Forms of Multiple Choice Testing

Author [redacted]
(Signature)

DAVID BRYAN ZANDVLIET
(Name)

September 29, 1993
(Date)