AN INVESTIGATION OF THE VALIDITY, RELIABILITY, AND ACCEPTANCE BY CHILDREN OF A MICROCOMPUTER ADMINISTRATION OF THE PEABODY PICTURE VOCABULARY TEST-REVISED

BY

TERRANCE GORDON LICHTENWALD

A dissertation submitted
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Psychology
California School of Professional Psychology
Fresno Campus
1986

CALIFORNIA SCHOOL OF PROFESSIONAL PSYCHOLOGY

FRESNO CAMPUS

The dissertation of Terrance Gordon Lichtenwald, "An Investigation of the Validity, Reliability, and Acceptance by Children of a Microcomputer Administration of the Peabody Picture Vocabulary Test-Revised," approved by his Committee, has been accepted and approved by the faculty of the California School of Professional Psychology, Fresno Campus, in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Psychology.

Dissertation Committee:

Norbert Ralph

Norbert Ralph, PhD

Ronald P. Unruh

Ronald Unruh, PhD

Alexander Gonzalez

Alexander Gonzalez, PhD
Committee Chairman

December 2, 1986

Mary Beth Kenkel

Mary Beth Kenkel, PhD
Dean for Academic Affairs

W. Gary Cannon

W. Gary Cannon, PhD
Campus Provost

DEDICATION

To my father Lloyd Phillip Lichtenwald who sacrificed so much for his family, I dedicate this work. It was through his support that I developed the desire and confidence to learn more about the world around me. It was his sense of humor which helped me to overcome the obstacles which blocked my road to success.

ACKNOWLEDGMENTS

This study was a lengthy, expensive endeavor. I wish to thank Rick Happ, President of Happ Electronics, Incorporated, for the financial support and technical expertise he offered and Sara Collins for her exceptional programming skill and artistic abilities. I am indebted to the consultants who worked on the project: Timothy Lichtenwald, school psychologist and systems analyst; Al Slatter, President of Software Experience, Incorporated; Sally Bunning, speech pathologist; and Lance F. Mahoney, speech pathologist/teacher of the learning handicapped.

My gratitude is expressed to each of my committee members: Alexander Gonzalez, for his willingness to chair a controversial dissertation and his professionalism; Norbert Ralph, for his knowledge of microcomputer systems and test administration with children and for his patience; Ronald Unruh, for his knowledge of test construction and current trends in education.

I wish to thank Joan Ensher and Inge Kauffman, who worked so diligently to obtain the books and articles related to this study.

I wish to thank the principal, teachers, parents, and children who willingly participated and made this study possible.

My appreciation is expressed for my closest friend, Chris Kuber, who has given me support and friendship through the years. I wish to thank Roman Welyczko for his wise legal counsel and friendship. I wish to recognize Dutch and Shirley Bindert, who have always provided me a safe port in any storm.

And to my mentor, Frances Cultertson, I wish to express gratitude for her encouragement and wisdom.

Most importantly, I thank my wife, Chris, for her love, understanding, patience, and faith. She has encouraged and motivated me to overcome the many obstacles which have stood in my way as I reach for my dreams. And I thank my mother and brothers for their unconditional love which has helped me through difficult and frustrating times.

Lastly, I would like to recognize all the learning-disabled individuals who have struggled as I have to overcome their handicaps and reach for their dreams.

ABSTRACT OF THE DISSERTATION

An Investigation of the Validity, Reliability, and Acceptance by Children of a Microcomputer Administration of the Peabody Picture Vocabulary Test-Revised

by

Terrance Gordon Lichtenwald

California School of Professional Psychology, Fresno Campus

Alexander Gonzalez, PhD

Dissertation Committee Chairman

1986

This study was an investigation of a microcomputer system which was developed to administer, score, analyze, and produce a written report for the Peabody Picture Vocabulary Test-Revised (PPVT-R). The variables of reliability, age, gender, level of previous experience with microcomputers, preference for manual or computerized test administration, and children's test performance under a manual and computer administration of the PPVT-R were evaluated.

The subjects in this study consisted of a cross-section of 98 children enrolled in Grades 2 through 6.

There was no significant discrepancy among the

test-retest reliability coefficients derived from the computer and manual test administrations. However, when the alternate form reliability coefficients derived from this study were compared to those reported in the PPVT-R Technical Supplement (Robertson & Eisenberg, 1981), a significant discrepancy was found for the fifth-grade computer-manual test sequence. The reliability coefficient was significantly lower than that reported in the PPVT-R Technical Supplement.

The subjects performed significantly better on the manual- than the computer-administered test, regardless of gender, age, or level of previous experience with microcomputers. But, those students who received the computerized administration first scored significantly lower than those students who received the reverse order of test sequence.

Those students exposed to microcomputer systems in their homes or classrooms performed significantly higher on both the manual and computer administration of the PPVT-R than those students who did not have computer systems in their homes or classrooms. Moreover, students with computers in their homes or classrooms performed significantly higher on the manual than the computer administration of the PPVT-R.

Analysis of the Test Preference Questionnaire found that 70.4% of the students expressed a preference for a computerized over human test administration, 20.4% reported no preference for either a human or computerized test administration, and 5% indicated that they preferred a human over a computer administration.

There were three limitations associated with the computer system used in the study. These were: (a) the heat generated by the computer, (b) the students' inability to alter their item selection under the computerized administration, and (c) the students' difficulty understanding some of the computerized stimulus words. Applications and implications for further study were indicated.

TABLE OF CONTENTS

Chapter | Page

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

The February 1986 issue of the American Psychological Association's Monitor contained 24 advertisements for computer hardware, software, or scoring and interpretation services having to do with the automation of different assessment instruments. H. J. Johnson (1984) has stated that:

> At the applied level, computerized testing has already had a major impact on the practice of psychology and psychiatry. More than 300,000 computerized test interpretations are processed annually by various scoring services. . . . There are now about 500 clinicians who either have terminals connected to or own prepackaged computerized assessment systems. . . . There are probably as many individuals who have programmed their own personal computer to administer and score psychological tests. (pp. 131-132)

Previously, computer applications in assessment typically involved the scoring and interpretation of test results. Recently, the interest in the development of tests which may be administered by microcomputers has increased. For example, National Computer Systems introduced the Psychometer, a microcomputer system which administered various psychological tests.

Although there have been numerous studies which support the reliability and validity of computer-administered self-report measures with adults, little is known about the effects of computer-administered testing with children.

There is little question that the development of a computer system which would reliably administer achievement tests to children at a reasonable cost would be valuable. Such a microcomputer system could serve as a screening instrument. It would not only administer the test in a reliable manner but would free the psychologist or teacher to spend more time working as a therapist, counselor, and educator instead of spending valuable time scoring and interpreting test data.

Given that computer-administered tests prove to be efficient, reliable, and accepted by children, new tests which take advantage of the features available with microcomputers could be developed. Computer-based tests could be designed to administer test items by utilizing three-dimensional graphics, multiple-color displays, and digital or synthetic voice. Moreover, the individual being tested could respond to the test items by using either a response console or through physiological responses. In the near future, it will also be possible for the individual being assessed to respond to the test items by giving a verbal response. An additional programming feature which could be developed would be software which is based on item analysis. For example, if the individual being assessed indicates on the test items that he is feeling depressed, the computer program would branch into the specific tests for depression. In addition, each individual assessed by a computer system could have his demographics and test scores added to the existing data base. Such data bases could be made available for numerous research projects.

In summary, the area of psychology regarding psychological evaluations has undergone a significant shift in the past decade. Currently, there is a plethora of computer programs which analyze and interpret standardized test batteries. In the last 10 years, there has been substantial interest in the development of systems which can administer test batteries.

This study contains the development of a microcomputer program for an Apple IIe computer which administers, scores, interprets, and produces a written report for the Peabody Picture Vocabulary Test-Revised. The variables of age, gender, previous computer experience, and preference for manual or computerized test administration was investigated to determine their effect on children's performance on an achievement test (PPVT-R) administered by a microcomputer.

Chapter 2

REVIEW OF THE LITERATURE

Presently, the use of microcomputers for the purposes of administration, scoring, and interpretation of psychological tests is a prevalent topic among psychologists. Moreland (1985) stated that:

> The use of computers in psychological assessment has recently become a hot topic, both within psychology and without. It is "hot" in the sense of giving rise to an increasing number of scholarly books and articles (e.g., Johnson, 1981; Schwartz, 1984), not to mention this journal. It is "hot" in the sense of giving rise to an ever increasing number of business enterprises (compare the advertisements in a recent APA Monitor with those in a 1981 issue). It is "hot" in the sense of capturing the attention of the news media (e.g., Hall, 1983; Petterson, 1983). And it is "hot" in the sense of giving rise to increasing controversy within the psychology profession. (p. 3)

The History of the Use of Computers in Psychology

The Development of Actuarial Systems

Fowler (1985) stated that "it is remarkable that the use of computers in psychological assessment is viewed as a new and revolutionary development" (p. 48). The author reported that computers began to be utilized in psychological testing almost immediately after they were available and that systems of computer-based test interpretation have been operational for almost 25 years.

Hedlund, Vieweg, and Cho (1985a) asserted that:

> one of the first uses of accounting machines in psychology and psychiatry was the development of a Hollerith card (punch card) form of the Minnesota Multiphasic Personality Inventory (MMPI) by Stark R. Hathaway and Kenneth E. Clark at the University of Minnesota in the early 1950s. (p. 5)

Although the Hathaway and Clark system marked a significant advance over the previous 15 years of hand scoring, the system required the services of a clinician for the translation of the scores obtained into descriptive statements. Hedlund et al. reported that the first actuarial computer system for interpreting the MMPI was developed at the Mayo Clinic in the early 1960s.

Although two of the computerized systems for the MMPI (the Institute of Living modification of the Mayo Clinic programs and the system developed by Gynther, Altman, & Sletten, 1973) permitted administration of the MMPI via the client sorting the Hollerith statement cards into a true and false pile, most MMPI scoring and interpretive systems utilized some type of MMPI booklet and machine-readable answer sheet which were completed by the patient.

Hence, throughout the 1960s and early 1970s, a psychologist was required to administer and score a psychological test (e.g., the MMPI) and mail the scores to a test center which owned the computer equipment necessary to interpret the test. Test results which were derived by such computer systems were called computer-based test interpretations (CBTIs). The MMPI supplied the input data for most CTBI systems, and these MMPI systems were the subject of most of the attempts to validate CBTIs.

With the advent of printer terminals and advances in telecommunications, it became possible for psychologists to send test data to computer centers over telephone lines. However, the psychologist was limited, since the data could only be entered when the center was able to accept the telephone call. Moreover, the psychologist was still faced with a time delay and the added expense of leasing the printer terminals in addition to the cost of the scoring service.

With the advent of microcomputer systems (e.g., Osborn, Apple, etc.) in the late 1970s, the necessity of sending the test data to a second party became obsolete. Specifically, the microcomputer system could interpret and produce its own CBTI for a psychological test via a computer "software" program which was contained on a 5¼-inch floppy disk.

The Development of Electromechanical Testing Systems

Despite a considerable amount of work that has been done regarding a wide range of automated testing, Hedlund, Vieweg, and Cho (1985b) noted that:

> The computer application which has received the most acceptance and use in the United States by psychologists has been associated with well-known self-report personality inventories (i.e., the Minnesota Multiphasic Peraonality Inventory [MMPI], the California Personality Inventory [CPI], the 16 Personality Factors [16PF], the Strong-Campbell Interest Inventory and the Minnesota Vocational Interest Inventory). (p. 5)

These early test administration systems involved complex electromechanical configurations of mechanical and electronic equipment (Birtles, Sambrooks, MacCulloch, & Holland, 1972; Elwood, 1969, 1972b, 1973; Elwood & Griffin, 1972; Geyde & Miller, 1970; Gilberstadt, Lushene, & Buegel, 1976; Klett & Pumroy, 1971; Klinge & Rodziewicz, 1976; Knights, Richardson, & McNarry, 1973; Overton & Scott, 1972; Paitich, 1973). Slide projectors or tape recorders with stimulus materials, special keyboards or response panels, and some type of electromechanical recording device, often integrated by paper tape, microswitches, or other electromechanical circuity, were utilized.

The Development of Microcomputer Test Administration Systems

Over the past 10 years, the complex electromechanical configurations have given way to microcomputer systems. A microcomputer-based system, the Psychometer, was typical of the microcomputer systems in use today. The Psychometer was a free-standing microcomputer which incorporated the software available thorugh the National Computer Systems, Inc. (NCS) telecommunication system. NCS reported that their microcomputer-based system was capable of administering the Beck Depression Inventory, Edwards's Personal Preference Schedule, Guildord-Simmerman Temperament Survey, Children's Personality Questionnaire, Jenkins Activity Survey, MMPI, Otis, Social History Questionnaire, and others.

Validity and Reliability of Automated Testing Systems

Several studies (Beaumont, 1985; Myers, 1981; Calvert & Waterfall, 1982; Elwood, 1969, 1972a, 1972c, 1973; Elwood & Griffin, 1972; Johnson, J. H., & Williams, 1980) have consistently demonstrated the reliability and essential comparability of test results between automated and manual administration of tests. Hedlund et al. (1985a) reported that free-standing microcomputer systems have been developed for the following tests: Bexley-Maudsley Automated Psychological Screening, Bexley-Maudsley Category Sorting Test, The Brief Intelligence Test, Continuous Performance Test, Digit-Digit Attention Test, Digit Span, Digit Symbol Substitution Test, Laterality Tests, Mill Hill Vocabulary Scale, Perceptual Maze Test, Raven's Progressive Matrices, Sbordone-Hall Memory Battery, Shipley Institute of Living Scale, and Psychological Systems Questionnaire. In addition, the Mental Health Treatment Service Computer System which was developed by the Salt Lake Veterans Administration currently holds 60 psychological and vocational tests and 20 histories and questionnaires in the computer system (Schwartz, 1984). The psychological programs are currently available to all Veterans Administrations Medical Centers.

Clearly, there has been a recent proliferation of microcomputer systems which administer numerous psychological tests. In addition, a variety of studies have examined the validity and reliability of automated testing. The published studies on automated testing have utilized adult subjects and found the reliability and validity to be acceptable (Dunn, T. G., Lushene, & O'Neil, 1972; Finney, 1966; Fowler, 1985; Lushene, 1981; Paitich, 1973).

Patients' Acceptance of Computerized Test Administration

In addition to the reliability and validity studies which investigated the effects of automated test administration, there have been numerous studies which have investigated patients' reactions to automated test administration. White (1983) found that 80% of college students preferred taking the MMPI by computer, while 20% expressed no preference. Carr and Ghosh (1983) found that phobic patients showed no apprehension about completing a full behavioral assessment by computer. Klingler, Johnson, and Williams (1976) found that 68% of the patients liked being tested on a computer, and 91% said they were as truthful and 45% said they were more truthful when tested by computer. J. H. Johnson and Williams (1980) found that 46% of their subjects said they were more truthful when responding to the computer than to a clinician. Angle and Carroll (1979) compared the responses of patients' self-reports of alcohol consumption under a computer versus a psychiatrist's interview. The researchers found that patients reported a significantly larger amount of alcohol consumption to the computer than to the psychiatrist. This finding suggests that some individuals may respond more honestly to computers.

To date, only two researchers have gathered data on children's acceptance of computerized test administration. Elwood and Clark (1978) reported that children judged computerized testing as "more like play than work," as "easy" instead of "hard," and judged the duration of a testing session as "short" as often as they judged it "long."

Katz and Dalby (1981) found that once children had been administered a test by computer, their perception of computers "improved." The researchers reported that the children's level of state anxiety showed a significant decrease between the first and second testing session irrespective of administration method or sample membership.

Computerized Test Administration With Children

There is a dearth of studies which address the variables associated with automated testing with children and adolescents. The following studies have addressed either the question of reliability and validity or children's acceptance of automated assessment.

Katz and Dalby (1981) investigated the variables associated with computer versus manual administration of a personality inventory with children. The authors administered the Fundamental Interpersonal Relations Orientation (FIRO-BC) and the State Trait Anxiety Inventory for Children (STAIC) questionnaires.

The subjects for the study were 40 gifted children and 40 behavior-problem children between the ages of 8 and 12. The gifted and behavior-problem children were combined and then divided into two groups. One group received a test-retest with a 1-week delay with the computer administration. The other group received the manual administration for both test administrations.

In both groups, the State Trait Anxiety Inventory for Children was administered prior to the Fundamental Interpersonal Relations Orientation. Previous to the administration of the questionnaires, the subjects completed an inventory designed by the authors to assess the extent of the subjects' previous experience and perceptions of computers. After completing the inventory, the group assigned to complete the questionnaires by computer received training in the use of the computer terminal.

Katz and Dalby (1981) conducted an analysis of their data in a three-step process. First, the reliability for both the experience with computers and the FIRO-BC were assessed. Second, the sample and treatment variances were analyzed. Finally, the effects of previous experience with computers on test performance were considered.

The researchers concluded that the computer questionnaire indicated generally favorable perceptions of computers by the subjects at the beginning of the study in addition to a significant test-retest stability. The analysis also found a significant test-retest correlation for the FIRO-BC. Katz and Dalby (1981) reported that the computerized method took significantly less time to administer and score than the traditional method.

A significant interaction for method of administration by testing session was also reported. Specifically, the interaction was attributable to changes in the children's perception of computers. This was viewed by the authors as support for the conclusion that subjects who received the computerized test significantly increased their positive perceptions of computers, while the perceptions of children using the traditional method did not change significantly. Katz and Dalby (1981) also reported that the subjects' levels of state anxiety showed a significant decrease from the first to second testing session, irrespective of the administration method or sample membership. The authors concluded that subjects who received the computer-administered test reported significant increases in positive perceptions of computers.

D. F. Johnson and Mihal (1973) published a study which addressed the performance of black and Caucasian children (seventh and eighth graders) across computerized and manual administration of the Cooperative School and College Ability Test (SCAT). A mixed design with the independent variables being method of administration (computer versus manual) and ethnicity of subject (black versus white) was employed. Ethnicity was the between factor, as each subject took both versions of the test, and the method of administration was the within factor. The method of administration was counterbalanced, so that each administration method occurred first in exactly half of the cases. The testing program ran with no time limit for item selection and no feedback regarding the correctness of the item selected.

The D. F. Johnson and Mihal (1973) study found a significant difference for black and Caucasian subjects on test performance. Specifically, the researchers reported a significant improvement in performance on the computerized test by the black students and no difference in performance as a function of test procedure for white subjects. A nonsignificant trend was reported for the white subjects.

Hence, both black and Caucasian subjects improved their scores with the computerized test. The study found that the Caucasian subjects scored significantly higher than blacks on the verbal and quantitative subtests of the SCAT under the manual administration. The Caucasian subjects also scored significantly higher than the black subjects on the quantitative subtest of the SCAT when it was administered by the computer.

D. F. Johnson and Mihal (1973) contended that the failure to find a significant difference between the method of administration for the white subjects was further evidence that computer administration was not detrimental to a subject's performance. The authors concluded that the finding of a significant difference in test performance of the black subjects under manual and automated administration pointed to several hypotheses which should be investigated further. First, the improved performance of the black subjects under the automated administration may stem from a decrease in the occurrence of a potential bias by the tester, both conscious and unconscious. Second, anxiety that may be readily induced when tests are administered to black children by persons representing more advantaged backgrounds was absent in this study. Third, the intrinsic motivation and interest in the testing situation for black subjects may be raised in the automated administration. Although D. F. Johnson and Mihal reported that it was not their intention to demonstrate which factors might be most problematic in minority testing, further research of the possible variables listed was called for.

Automation of the Peabody Picture Vocabulary Test

Four of the articles reviewed address the findings of past researchers who automated the Peabody Picture Vocabulary Test (PPVT). The PPVT is an adaptable test to computerize due to its totally visual nature and the lack of necessity for the examiner to interact with the subject other than to present the stimulus words. During the administration of the PPVT, the subject is required to match a correct word with a picture which expresses the concept of the word. For example, the subject must select from four pictures (e.g., ball hat, bat, wagon) the pictured object which is represented by the word "ball."

Klinge and Rodziewicz (1976) studied the advantages of automated testing with the Peabody Picture Vocabulary Test (PPVT). Both the automated and manual testing was conducted by a single administrator. The subjects were assigned to one of two conditions. One group received the automated administration first, and then, the manual. The second group received a reversed order of test administration. No significant differences in test version, order of administration, or gender were reported.

The researchers also raised the question of the strength of their correlations between the manual and automated test administration. Specifically, although the correlation analysis showed significant correlations, the correlations in the study were lower than other test-retest administrations of the PPVT. Two previous studies conducted by Overton and Scott (1972) and Knights et al. (1973) used retarded subjects. These studies found higher correlations.

Klinge and Rodziewicz (1976) developed two hypotheses to account for the lower correlations in their study. First, the subjects utilized in this study were 52 psychiatrically disturbed adolescent inpatients. The lower correlations could be due to a tendency for the subjects' behavior to be varied. Consequently, their test performance would be less stable across time. A second hypothesis was that the subjects might have been differentially intrigued by the automated administration. The authors stated that this would be an interesting area of future research.

Klinge and Rodziewicz (1976) concluded that the advantages of the automated version of the PPVT included the following: (a) high reliability between the manual and automated versions, (b) reduction in testing time, (c) increased cooperation of the subjects in the automated sessions, (d) freedom of the administrator to observe the subjects' behavior, (e) immediate availability of test data given by the computer scoring, and (f) increased standardization of both administration and scoring.

The focus of a study conducted by Knights et al. (1973) was to study the feasibility and reliability of an automated administration of the Peabody Picture Vocabulary Test (PPVT) and the Raven Colored Progressive Matrices (CPM). The subjects were divided into two groups. One group received the automated, and then, the manual administration of the PPVT and CPM. The second group received the reversed order of test administration. Under all testing conditions, the PPVT was administered before the CPM. All tests were administered by the same individual, who was also present in the room with each child during the automated testing. The test administration instructions were altered. The words "point to" and "show me" were replaced by "touch."

Knights et al. (1973) utilized a two-way analysis of variance (condition x testing) with repeated measures on the last factor for each of the two tests. The analysis yielded no significant main effect for condition or testing with the PPVT. However, a condition x testing interaction was significant. Hence, the analysis indicated that those subjects who first "performed the automated procedure scored lowest and improved more on the second testing than those who were first tested clinically" (p. 224). This finding held for both the PPVT and the CPM.

It was the researchers' contention that the reason for the significant difference between the order of test administration was due to the exposure of the subjects to a new situation and the difficulty the subjects had in adapting to it. The authors supported their hypothesis of environmental influences by pointing out that the Overton and Scott (1972) study also found differences in the ability of retarded children to respond to the automated procedure.

The test-retest correlation coefficients for the PPVT and CPM were significant and determined by Knights et al. (1973) to compare favorably with reliabilities reported in the test manuals. It was the authors' contention that the lower correlations for the CPM were related to the difficulty the subjects had in understanding the demonstration item with the automated procedure. The researchers reported that the automated test administration took longer than the manual. However, the study found that the automated scoring was "much more rapid" than the manual. The total administration for the two procedures was about the same. The authors also reported that the subjects enjoyed interacting with the automated administration.

The study designed by Overton and Scott (1972) studied the correlations of scores on both forms of the PPVT under manual and automated presentation in an immediate test-retest paradigm with the orders of testing counterbalanced. The subjects in the study were 240 retarded young adolescents.

The Overton and Scott (1972) study found no apparent warm-up or practice effects during either the manual or automated presentation. The authors reported a significant difference between the grand mean for first and second administration. A difference between the parallel forms (Form A and Form B) of the PPVT was also found. The researchers concluded that the Form B was more difficult than Form A. The discrepancy between Forms A and B was also evident under the automated administration. However, the difference was not statistically significant. The correlations found under the automated form were as high as the correlations under the manual administration.

In terms of administration time, the Overton and Scott (1972) study found that the automated administration took longer than the manual administration. It was not surprising that the automated method took longer, given that the authors had used a fixed interitem interval of 2 seconds for the automated method. The authors concluded that the fixed interitem interval could be adjusted to other time ratios. The 2-second interval was an arbitrary choice.

Another significant finding of the Overton and Scott (1972) study was the number of subjects who did not reach the basal level on the PPVT. Specifically, there were more subjects who were unable to get a basal score with the automated than with the manual version of the test. The authors interpreted this finding to indicate that the absolute rates of failure were low and that the disadvantage of the automated version would most probably be eliminated if a minimal level of pretraining was extended for the subjects.

Overton and Scott (1972) concluded that the above study "clearly demonstrated the possibility of using automated tests with retarded subjects" (p. 642). In addition, the authors contended that, once the technological problems of the administration were eliminated, the use of automated testing would become widely available. Further research on automated testing was suggested, due to the initial failure rates with the automated equipment. It was suggested that careful training procedures could replace the crude methods used in the present study. Another important finding of the Overton and Scott study was the difference between Forms A and B of the PPVT. It was recommended that the differences between Forms A and B be noted carefully by individuals using the PPVT in any test-retest model or where gain scores are of interest.

The study conducted by Elwood and Clark (1978) was an investigation of the statistical properties and practical clinical aspects of the Peabody Picture Vocabulary Tests (PPVT) when administered via an electromechanical system. Every subject began with the three sample items, followed by Item 1 of the test and continued until reaching the discontinuation criterion. Following each test, the administrator asked the subject if the test had been "more like work" or "more like play," if it had been "easy" or "hard," and "long" or "short." The study utilized 70 children between the ages of 4 and 13 years of age.

One of the findings of Elwood and Clark (1978) was that male subjects obtained significantly higher IQs than female subjects. The authors had difficulty accounting for the IQ differences between the sexes. Elwood and Clark concluded that some experimental conditions inhibited the performances of female subjects. It was suggested that further research on computerized testing of young children may be needed to control the influences of gender on stimulus words and the gender of the test administrator.

In addition, the Elwood and Clark (1978) study did not find the test-retest correlations at the level of previous studies. The researchers felt that the different correlations reported by the different studies raised questions about the comparability of different automated systems and the influences of subjects' age on test-retest reliability. Specifically, the Knights et al. (1973) and Overton and Scott (1972) studies utilized touch sensitive panels; whereas, the response required of the subjects in the Elwood and Clark study involved asking subjects to look at the screen, select the picture that matched the stimulus word, determine the number associated with the picture, select the push button with a comparable number, and then, press the push button. The authors hypothesized that the response method utilized in their study may have required a higher level of functioning than that of the Knights et al. and Overton and Scott studies.

Elwood and Clark (1978) found that children judged computerized testing as "more like play than work," as "easy" instead of "hard," and judged testing sessions duration as "short" as often as they judged them as "long." In addition, the subjects' request for help decreased as they gained experience with the computerized system, and the time they spent studying test items increased as the items became more difficult. Finally, the IQ estimates obtained using the computer system were essentially the same between first and second sessions, between Form A and Form B, and between older and younger subjects.

In summary, the Peabody Picture Vocabulary Test has been administered in an automated manner in four studies (Elwood & Clark, 1978; Klinge & Rodziewicz, 1976; Knights et al., 1973; Overton & Scott, 1972). However, the method of automation in each study was different. Each author used his own interpretation of the concept of automation. For example, within the Overton and Scott (1972) study, the subjects were exposed to the verbal directions of a test administrator via an intercom system during the automated phase. Within the Klinge and Rodziewicz (1976) study, the subjects were given the stimulus words via a tape recording (female voice). Throughout each of the studies, the lack of a standard method of automation made it impossible for the studies to be truly compared and contrasted.

Given the lack of standardization, it is surprising that the studies yield similar findings in several aspects. For example, in each study significant correlations were found between the manual and automated method of test administration. However, in the Klinge and Rodziewicz (1976) study, the authors questioned the strength of their correlations. The Knights et al. (1973) study was the only one which found a significant condition x testing interaction. Specifically, those subjects exposed to the automated test administration first scored lowest and improved more under the manual test administration than those who were first tested manually. This study contributed the significant interaction to characteristics of the subjects utilized in the study.

The explanation submitted by Knights et al. to account for the condition x testing interaction was also addressed by each of the authors of the other studies in their discussion sections as a rationale for the correlation trends observed. Klinge and Rodziewicz (1976) attributed their correlation findings to the fact that they utilized an adolescent psychiatric population whose behavior was less stable across time, and thus inconsistent under testing conditions. Overton and Scott (1972) were also forced to address a similar condition, since there were several subjects who could not reach the basal level on the PPVT during the automated administration. They contributed the subjects' difficulties under the automated condition to the testing equipment and lack of additional test illustrations.

A specific area of concern in automated assessment was the finding that children with different disorders appear to have different levels of adjustment to the testing environment and performance ability. A second significant finding was that the automated method did not decrease the testing time. Specifically, one of the hypothesized advantages of automated administration was the time saved in testing. Each of the studies did not find a significant savings in the time required for administration but did report a significant savings in time spent in scoring the test. A third finding was the poor economic feasibility of automated test administration. In terms of cost effectiveness, a school/clinical psychologist, in an applied practice, would find the cost of purchasing all the equipment necessary for the administration of the PPVT prohibitive. In addition, the purchase of such equipment would be a poor investment because it would only be used to administer the PPVT.

There is little question that the investigations conducted by Overton and Scott (1972), Knights et al. (1973), Klinge and Rodziewicz (1976), and Elwood and Clark

(1978) have served as a starting point for the study of the effects of computerized testing on children's performance. However, a significant factor has changed since the early automated PPVT studies. Specifically, the average school-age child today has had far more experience with the technology of computer science than his predecessors.

The Use of Microcomputers in Education

Kulik, Kulik, and Bangert-Drowns (1985) reported that "computers are fast becoming an important factor in elementary school teaching. The number of computers in American elementary schools has increased by a factor of at least 10 during this decade, and most schools now own them" (p. 59). The Kulik et al. study was a meta-analysis of 32 comparative studies. The authors concluded that computer-based education has generally had positive effects on the achievement of elementary school pupils. Kulik et al. divided the computer instruction into off-line computer-managed instruction (CMI) and interaction computer-assisted instruction (CAI). The study found the average effect in 28 studies of CAI programs was an increase in pupil achievement scores of 0.47 standard deviations, or from the 50th to the 68th percentile. However, the average effect in four studies of CMI programs was an increase in scores of only 0.07 standard deviations.

Although there are numerous studies such as those utilized in the Kulik et al. (1985) study which support the effectiveness of interactive computer-assisted instruction programs with children, there are no studies which control for the demographic and affective factors. These factors might affect the child's ability to intearct with a computer-based instruction or testing system. Therefore, although there are 28 studies which support the "teaching" of children via interactive computer systems, there is a need for a study to investigate the effects of testing children via interactive computer systems.

Conclusion

This review of the literature addressed the history of computers in psychology. The literature indicated that there has been a shift in the focus of the utilization of computer-based systems in the applied psychology field. As early as the 1950s, mainframe computer systems were being employed to run actuarial programs for psychological tests, such as the MMPI. With the technological advances in computer science, the ability of computers to assist psychologists with scoring, and then, interpretation of psychological tests became evident. As the level of the psychologist's ability to score and interpret psychological tests via computers grew, there were numerous studies which investigated the reliability and validity of computer-based interpretations. With the development of telecommunication systems, the era of the psychologist having tests administered, scored, and interpreted by computer became an economic reality. As microcomputer systems developed, the interest in developing systems which administered and interpreted psychological tests led to the formation of corporations which marketed such microcomputer systems.

There have been many studies which found the validity and reliability of psychological tests administered by computer to be acceptable, and in some instances, computer-administered interviews are preferable. However, there are few studies which investigated the validity and reliability of computer-based testing systems with children. The early studies which researched the effects of automated administration with children utilized primarily the PPVT, CSAT, and FIRO-BC. These early studies reported different findings for automated versus manual administration. The discrepancy in the findings for the PPVT was due, in part, to the different electromechanical systems developed by each researcher. In addition, the electromechanical systems did not follow the standardized rules for test administration due to the limited computer technology which was available.

The present study was an investigation of the variables which have been reported to effect children's ability to be assessed by automated methods. However, the present study can be differentiated from the earlier studies which used the PPVT. The present study employed a self-contained microcomputer system (i.e., Apple IIe). It administered, scored, interpreted, and produced a written report for the PPVT-R without an external electromechanical apparatus. In addition, the present study investigated the variables of age, gender, previous computer experience, and preference for manual or computerized administration on children's test performance.

Research Hypotheses

Hypothesis 1

It is hypothesized that the test-retest correlation coefficients for the computer versus manual test administration of the PPVT-R will be significantly different.

Hypothesis 2

It is hypothesized that there will be a significant difference between the test-retest correlations for Forms L and M reported in the PPVT-R Technical Supplement and those derived in this study.

Hypothesis 3

It is hypothesized that there will be a significant discrepancy between the order of test administration (i.e., computer versus manual and manual versus computer) and test-retest reliability coefficients.

Hypothesis 4

It is hypothesized that there will be a significant difference between the subjects' preference for the computerized versus manual test administration.

Hypothesis 5

It is hypothesized that (a) subjects who have microcomputer systems in their classroom will score significantly higher on both the manual and computer test administration than those subjects who do not have computer systems in their classrooms, and (b) subjects who have microcomputers in their homes will score significantly higher on both the manual and computer test administration than those subjects who do not have computer systems in their homes.

Hypothesis 6

It is hypothesized that there will be a significant difference between male and female test performance under the computer and manual test administration.

Chapter 3

METHOD

Subjects

The subjects for the study were selected from second- through sixth-grade students at an elementary school within the Fresno Unified School District in Fresno, California. The classroom teachers for each of the five grades gave each student in their classrooms a letter addressed to his or her parents describing the research project and a permission slip (see Appendix A). The parents were requested to sign and return the permission slips.

Initially, the study consisted of a total of 125 students with 25 students being chosen from each grade. Of the 125 subjects who were selected to participate in the study, 27 were dropped from the study. All the subjects who were dropped were students who were not released from their classrooms to participate in the second test administration. Thus, the study consisted of 18 subjects from Grade 2, 20 from Grade 3, 20 from Grade 4, 21 from Grade 5, and 19 from Grade 6, for a total of 98 subjects. Forty-five of the subjects were male and 53 were female.

Design

The subjects were randomly placed into either a sequence of manual administration, followed by the computer administration, or a sequence of computer administration, followed by the manual administration, of the PPVT-R. The interval between the test-retest for the PPVT-R was 1 week. Form L of the Peabody Picture Vocabulary Test-Revised (PPVT-R) was administered by computer. In the manual test condition, Form M of the PPVT-R was given. During the first administration of the PPVT-R test sequence, the subjects filled out an inventory assessing their previous experience with microcomputers (see Appendix B). After they had been exposed to both the computer and manual testing conditions, the subjects filled out a questionnaire to determine which administration condition they preferred (Appendix C).

Two female students who were enrolled at California State University, Fresno, were recruited to administer the PPVT-R. The test examiners were assigned to either the computer administration or the manual administration condition via a random draw. The test examiner's placement into a test sequence was then counterbalanced. Each examiner was trained in the administration procedures for her assigned test administration condition as outlined in the Procedure section. Neither test administrator was informed of the hypotheses of the study.

Procedure

A consent from was received for each subject prior to administering the PPVT-R (see Appendix A). The legal guardians of other potential subjects either did not receive the request letters and permission forms from their children or returned the request forms with their refusal to allow their children to be tested. In addition, prior to the first test administration, each subject completed a questionnaire designed to assess his previous experience with microcomputers (see Appendix B). Each classroom teacher who had students who were participating in the study received a subject list.

Subjects were released from the classroom 2 at a time. Both the computer and manual testing administration areas were located in separate areas of the school library. One subject went to one area for the computer-administered condition, and the second subject went to another area for the manual-administered condition. The following week, the subjects went to the opposite testing area to be given the counter condition.

Each testing area had two chairs, one of student size and one of adult size, and a table (or flat-top desk) designed for elementary school children. The seating arrangement for the manual administration condition followed the seating arrangement recommended by L. Dunn and L. Dunn (1981). Specifically, the examiner and subject were seated around the corner of the table. The easel containing the series of test plates was placed so that the subject could see only the plate being administered. The Individual

Test Record Form M was placed behind the easel to shield the form from the subject's view. For left-handed examiners, the seating arrangement was reversed.

For the computer administration, the subject was seated directly in front of the Apple IIe, response console, and display monitor. The subject was seated so that he could reach the microcomputer keyboard and response console and view the monitor at eye level. The subject was informed that the computer would administer the directions to him or her. In addition, the subject was informed that if he or she would like the directions or a word repeated, he or she should press the space bar on the computer keyboard.

A subject under the chronological age of 8 was shown three PPVT-R training picture plates (A, B, and C) to introduce the test format and the examiner's verbal instructions. Training plates D and E were given to each subject 8 years and older. After the training plates were administered, the subject was shown the picture plate corresponding to the item number that is recommended as a starting point on the Individual Test Record form. The starting points are for subjects assumed to be of average ability (Dunn, L., & Dunn, L., 1981). Each subject obtained a basal of 8 consecutively correct responses and a ceiling of 8 consecutive responses containing 6 errors.

The picture plates administered, the verbal instructions given during the test, and the method of determining the basal and ceiling levels were identical under the manual and computerized administration with one exception. Under the computerized administration, the verbal instruction, "point to" was replaced with "push the correct button for" the picture which best tells the meaning of the word. If, during the computer administration of the test plates, the subject did not push the correct button for the corresponding image on the plate, the computer informed the subject that he had made a good try and told him the correct number he should have picked. The computer program followed the format for incorrect responses as outlined by L. Dunn and L. Dunn (1981). If a subject experienced difficulty in either the manual- or computer-administered condition with the selection of the appropriate image, the test administrator followed the guidelines described by L. Dunn and L. Dunn (1981). After the subject had completed both testing conditions, he completed the questionnaire regarding his preference for the manual- or computer-administered tests (see Appendix C).

A feature which distinguished the manual from the computerized administration of the PPVT-R was the scoring of the Individual Test Record form. Once the subject had been dismissed from the testing room under the computerized administration condition, his protocol was recorded, scored, and interpreted. A report was produced by the computer before the next subject was assessed. It was necessary for the computerized system to run the subject's protocol as the program was not designed to store data files.

Measures

The study utilized two self-report inventories: the Previous Experience With Microcomputers Inventory (PEMI) and the Test Administration Preference Questionnaire (TAPQ).

Previous Experience With Microcomputers Inventory

The Previous Experience with Microcomputers Inventory (see Appendix B) was developed to measure the extent of a subject's previous experience with microcomputers. The inventory assessed the type of computer system the individual had been exposed to in the home, school, and community. The inventory also measured the amount of exposure and the subject's attitudes toward computers.

Test Administration Preference Questionnaire

The Test Administration Preference Questionnaire was developed by this writer (see Appendix C). The Test Administration Preference Questionnaire consisted of questions regarding time perception (i.e., Which took longer? When the computer or the person gave you the test?), degree of difficulty (i.e., Which type of test was harder?

The one given by the computer or the person?), and motivation level (i.e., Which was more like play or work? The test given by the computer or the test given by the person? If you were to take a test again, would you like to take it on a computer or have someone administer it to you?).

The Peabody Picture Vocabulary Test-Revised

The Peabody Picture Vocabulary Test-Revised (PPVT-R) was used in this study as the basis for the comparison between computer and manual test administration. The PPVT-R is an individually administered, norm-referenced, wide-range, power test of receptive vocabulary for individuals between the ages of 2.5 through 40 years.

There are two parallel forms (L and M) of the PPVT-R. This study employed the use of Form L for the computer administration and Form M for the manual administration. Both Forms L and M contain five training items, followed by 175 test items arranged in order of increasing difficulty. Each item has four black-and-white illustrations arranged in a multiple-choice format. The subject's task is to select the picture considered to best illustrate the meaning of a stimulus word presented orally by the examiner. The starting point on the test is determined according to the subject's chronological age. The basal is established when the subject makes eight or more consecutive correct responses. The ceiling is the lowest eight consecutive responses containing six errors.

The PPVT-R uses a mean of 100 and standard deviation of 15. Based on the chronological age of the client and the raw score, the examiner can obtain three deviation-type age norms which are contained in the test record: standard-score equivalents, percentile ranks, and stanines. In addition, development-type age norms and age equivalents can also be reported. The original PPVT produced a deviation intelligence quotient index. However, the PPVT-R has replaced the term deviation intelligence quotient with the more neutral term "standard score" (Dunn, L., & Dunn, L., 1981, p. 91).

The PPVT-R was standardized on a national sample of 5,028 individuals (4,200 of the normative sample were children and adolescents and 828 were adults). The PPVT-R derived its normative data by administration of either Form L or Form M of the PPVT-R to a subject. L. Dunn and L. Dunn (1981) selected the one form method of standardization because they felt that administering only one form eliminated the practice effect problem that would have resulted had both forms been administered to each subject. In addition, the authors felt that the one form standardization method would yield a normative sample with twice as many subjects.

The reliability coefficients reported for the PPVT-R by Robertson and Eisenberg (1981) addressed internal consistency (split-half), delayed retest alternate-forms, and practice effects.

The split-half reliability coefficients were calculated via the Rasch-Wright model. The Rasch-Wright procedure for calculating split-half reliability was selected because the standard procedure for calculating split-half correlations produced spuriously high coefficients. The spurious coefficients occurred because items below the basal were counted as correct, and items above the ceiling were counted as incorrect. The Rasch-Wright model considers only the items answered correctly between the basal and ceiling items in addition to item difficulty. Once these data were derived, the resulting split-half correlations were corrected for half-test length by the Spearman-Brown formula. The corrected split-half reliabilities for Form L yielded a range from 0.67 to 0.88, with a median of 0.80 for the ages 2.5 through 18 years old. The corrected split-half reliabilities for Form M yielded a range from 0.61 to 0.86, with a median of 0.81 for the ages 2.5 through 18 years old. The adult standardization split-half coefficients were reported for Form L but not for Form M. The adult split-half coefficients for Form L ranged from a 0.80 to 0.83, with a median of 0.82.

The delayed alternate-forms retest reliability coefficients for raw scores and standard score equivalents were reported by both age and grade level. The delayed retest reliabilities based on raw scores for separate grade groups, preschool through Grade 12, ranged from 0.47 to 0.91, with a median of 0.76. The delayed retest reliabilities based on raw score for age groups ranged from 0.52 to 0.90, with a median of 0.78. The delayed alternate-forms retest reliabilities based on standard score equivalents for age groups ranged from 0.54 to 0.90, with a median of 0.77.

Practice effects based upon a delayed retest averaged less than 1 standard score point for all age groups. Overall, the practice effects for the PPVT-R appeared to be inconsequential.

Design of the Computer Program

The program was written in Beginner's All-Purpose Symbolic Instruction Code (BASIC) version Applesoft. The computer program was written by Sara Collins for Happ Electronics, Incorporated. The Applesoft version of BASIC was selected over Integer BASIC due to Applesoft's foundation in the Apple IIe firmware. Specifically, Applesoft BASIC allowed the programmer the option of direct commands to access the input/output ports. In addition, Applesoft offered superior string commands and flowing decimal points.

The Mountain Computer Speech Digitizer (MCSD) utilized in this study followed the basic format of the microprocessor developed by Davidheiser (1982). Given the program control variables offered with digital voice and the MCSD, specifically, digital voice was chosen over a synthesized speech program.

The PPVT-R version 1.0 speech program was an exact reproduction of the instructions given in the PPVT-R Forms L and M Manual (Dunn, L, & Dunn, L, 1981) except that the command "point to" was replaced with "push the correct button for."

A Digital Visual System was employed to develop the high resolution graphics used to display the picture plates for the PPVT-R version 1.0. The images were then enhanced via a light pen system developed by Magellan Light Pen Systems (1985) which boosted the high resolution graphics into double high resolution. Once the images were developed into double high resolution, the Digital Visual System and Magellan Light Pen System required a conversion process routine to use the double high resolution on the Apple II system.

The PPVT-R computer program was designed to score and calculate the descriptive scores which can be derived from the PPVT-R. Specifically, the program calculated a subject's chronological age, raw score, standard score, percentile rank, stanine, confidence band for age equivalents with the obtained test scores and lower and upper band width, and standard error of measurement for raw scores--median value. In the PPVT-R test report there was a list of all items administered, including which items were correctly or incorrectly identified by the subject. Moreover, the PPVT-R report listed the subject's basal and ceiling items.

Chapter 4

RESULTS

Hypothesis 1--Analysis of Alternate-Form Reliability Coefficients

The alternate-form reliability coefficient for each grade, based on the raw scores, was computed and then corrected for differences in the variance of test Forms L and M as suggested by Angoff (1953). Each pair of correlation coefficients was tested for comparability, using a test of significance of independent correlation (see Table 1).

None of the pairs reached statistical significance. One possible reason for the failure to find a significant discrepancy between the correlation coefficients stemmed from the small number of students used in the study. Specifically, the small number of students may have resulted in unstable correlation coefficients. This, in turn, resulted in a lower probability of detecting significant discrepancies.

A two-sample independent Z test (McCall, 1980) was used to examine the difference between the test sequence. The analysis did not find a significant statistical difference in the comparison of the corrected correlated

Table 1

Alternate-Forms Reliability Coefficients by Grade for Raw Scores

			Raw score				t test		
			Form L		Form M				
Grade	N	Testing group	Mean	SD	Mean	SD	r	p	Correct
2	9	L-M	77.89	15.02	85.78	13.59	.789	.036	.798
	9	M-L	90.11	11.48	91.67	9.55	.925	ns	.932
3	10	L-M	92.80	8.65	101.60	13.58	.203	.088	.151
	10	M-L	103.90	8.84	104.40	9.41	.757	ns	.750
4	11	L-M	98.91	24.43	101.27	15.06	.881	ns	.908
	9	M-L	105.44	16.85	108.78	12.97	.912	ns	.923
5	11	L-M	115.45	9.15	120.09	15.08	.228	ns	.167
	10	M-L	114.70	13.80	114.60	16.28	.934	ns	.929
6	8	L-M	113.00	14.77	115.78	14.86	.844	ns	.844
	10	M-L	115.20	8.92	120.70	9.17	.369	.121	.364

scores within each grade level of the testing sequences L-M or M-L.

Hypothesis 2--Analysis of Correlations Derived in This Study and Corresponding Correlations Reported in the Technical Supplement

The corrected correlations found in this study for the L-M and M-L test sequence by grade level, based on the raw scores, were compared to the corresponding correlations reported in the Technical Supplement for the PPVT-R via a two-sample independent Z test (see Table 2). The only statistically significant difference occurred in the fifth-grade L-M test sequence. The L-M test sequence found in this study was significantly lower than that reported in the Technical Supplement (r corr = 0.167 and r corr = 0.94, respectively).

Hypothesis 3--Analysis of Order of Test Administration

A three-way analysis of variance (grade level x test sequence x test format--computer or manual) was conducted to assess the influence of test sequence and test format on test performance (see Table 3). As presented in Table 3, test format reached statistical significance and test sequence was marginally significant. The absence of any interaction effect warranted the interpretation of the two main effects. Within the test sequence condition, those students who took the manual format first and the computer

Table 2

Alternate-Forms Reliability Coefficients From Normative and Derived Data

Grade	Testing group	r from normative data	r from derived data	t test p
2	L-M	.56	.78	ns
	M-L	.67	.93	ns
3	L-M	.78	.15	ns
	M-L	.74	.75	ns
4	L-M	.80	.90	ns
	M-L	.79	.92	ns
5	L-M	.94	.16	$(p < .05)$
	M-L	.82	.92	ns
6	L-M	.91	.84	ns
	M-L	.76	.36	ns

Table 3

Three-Way Analysis of Variance (Grade Level x Test Sequence x Test Format) Summary Table

Source	SS	df	MS	F	p
Between subjects		97			
Grade	1,923.87	4	480.97	1.038	
Testing sequence	1,925.53	1	1,925.53	4.156	(.045)
Grade x test sequence	1,904.42	4	476.10	1.027	
Error between	40,776.29	88	463.37		
Within subjects		98			
(FORMAT)	1,255.18	1	1,255.18	15.010	(.001)
Grade x format	37.38	4	9.35	0.112	
Test x format	139.53	1	139.53	1.669	
Grade x test x format	242.28	4	60.57	0.724	
Error within	7,358.62	88	83.62		
Total		195			

format second did better under each test administration than those students who took the test administration in reverse order, $F(1, 88) = 4.156$, $p < .045$. Within both test sequence conditions, students scored significantly higher on the manual test administration, $F(1, 88) = 15.010$, $p < .001$. As illustrated in Figure 1, those students who took the manual followed by the computer test sequence (M-L) did better than the students who took the opposite test sequence (L-M). In addition, the students' manual test scores were higher than the computer test scores.

Hypothesis 4--Analysis of the Test Administration Preference Questionnaire

With respect to test administration preference, 69 of the 98 students (70.4%) surveyed, reported that they would prefer to take their future examinations from a computer rather than having the test administered to them by a human. In addition, 24 students (24.4%) reported no preference for either a computer or a human test administrator, and 5 students (5.10%) reported that they would prefer a human test administrator over a computer for their future tests. Due to the skewed distribution of the students' preference, further analysis of test preference and test performance is unwarranted.

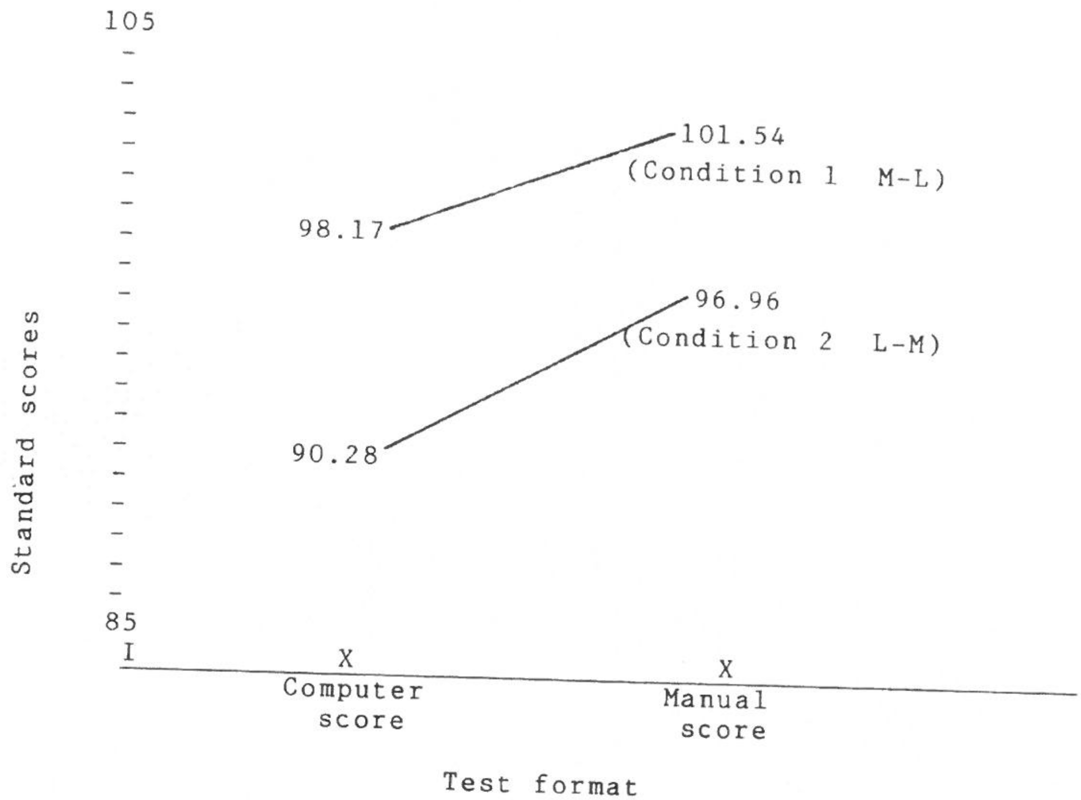

Figure 1. The influence of test sequence and test format on test performance.

Hypothesis 5--Analysis of the Previous Experience With Microcomputers Inventory

The Previous Experience with Microcomputers Inventory found that 64 of the 98 students surveyed (65.3%) reported having a microcomputer system in their home. In terms of the students' experience with computers in their classroom, 6 (6.1%) reported having a computer in their classroom, 28 (28.5%) reported that there was occasionally a computer in their classroom, and 64 (65.3%) reported no computer system in their classroom. Eighty-nine of the students (90.8%) reported having gone to video game rooms to play video games, and 9 (9.1%) reported that they had never played a video game in a game room.

A three-way analysis of variance (computer in the classroom x test sequence--L-M or M-L x test format--computer versus manual) was conducted to assess the effect of having a computer in the classroom on test performance (see Table 4).

As presented in Table 4, two main effects (experience and test sequence) reached statistical significance. The effects of test sequence and interaction between experience and test sequence were marginally significant. The interaction effect was caused by the low scores marked by the testing sequence (manual-computer) for students who had no computer at school. Their scores were markedly lower in both computer and manual format (82.53 and 89.65)

Table 4

Three-Way Analysis of Variance (Computer in the Classroom x Test Sequence x Test Format) Summary Table

Source	SS	df	MS	F	p
Between subjects		97			
Experience	1,785.97	1	1,785.97	4.026	0.047
Test sequence	1,572.74	1	1,572.74	3.546	0.062
Experience x test sequence	1,477.33	1	1,477.33	3.331	0.071
Error between	41,694.07	94	443.55		
Within subject		98			
(FORMAT)	1,255.18	1	1,255.18	15.497	0.001
Experience x format	1.48	1	1.48	0.018	
Test x format	137.88	1	137.88	1.702	
Experience x test x format	24.98	1	24.98	0.308	
Error within	7,613.47	94	80.99		
Total		195			

than other conditions. The marginally significant interaction somehow obscured the interpretation of the main effect of the experience. However, those who have access to a computer at school did better than those who did not have access to a computer at school (see Figure 2). The statistically significant test format effect, $F(1, 94) = 15.50$, $p < .001$, shows that scores on the manual format were better than the scores on the computer format ($\overline{X} = 99.20$ and 94.14).

Those students who have access to a computer in their school did significantly better on both the manual- and computer-administered test than those students who did not have access to a computer in their school (Figure 2). An additional finding was the influence of test sequence on test performance. Students who did not have access to a computer in their school were at a disadvantage in the computer/manual (L-M) test sequence on both the computer and manual test format (see Figure 3).

A three-way analysis of variance (computer at home x test sequence L-M or M-L x test score) was conducted to assess the effect of having a computer in the home on test performance (see Table 5). Two main effects (experience and test format) reached the statistical significance. No interaction effect was statistically significant. The students scored higher on the manual test format than the computer test format. In both test formats, the students

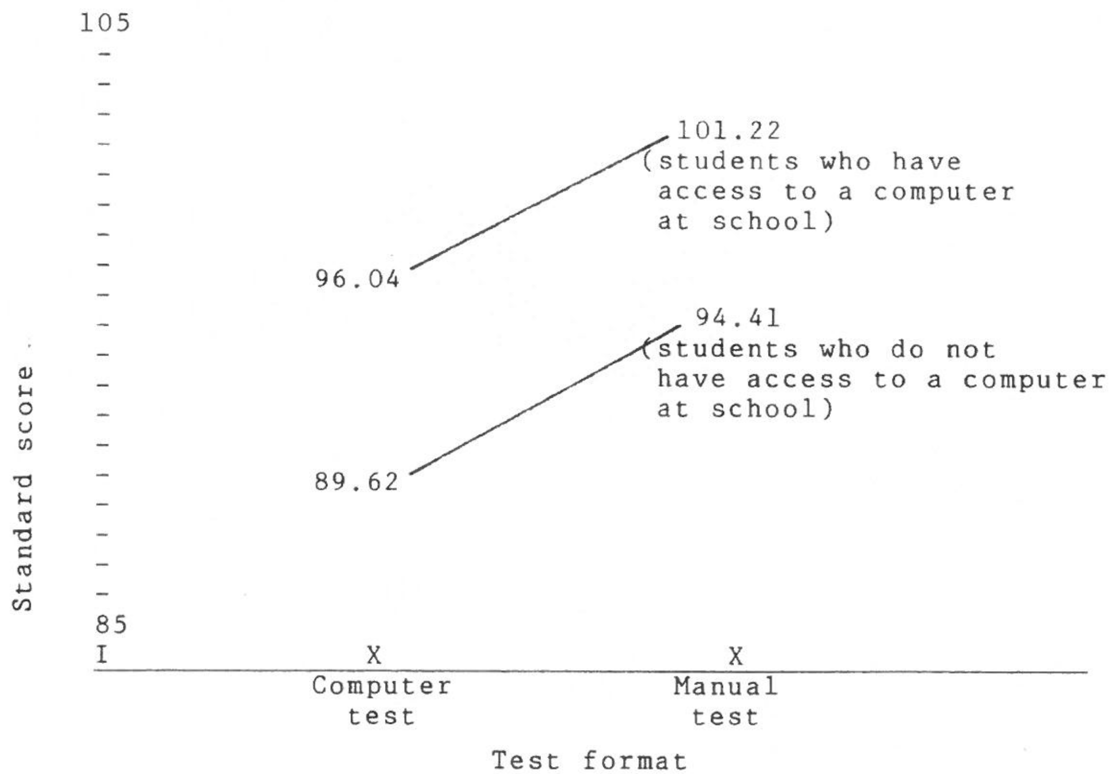

Figure 2. The influence of access to a computer at school and test format on test performance.

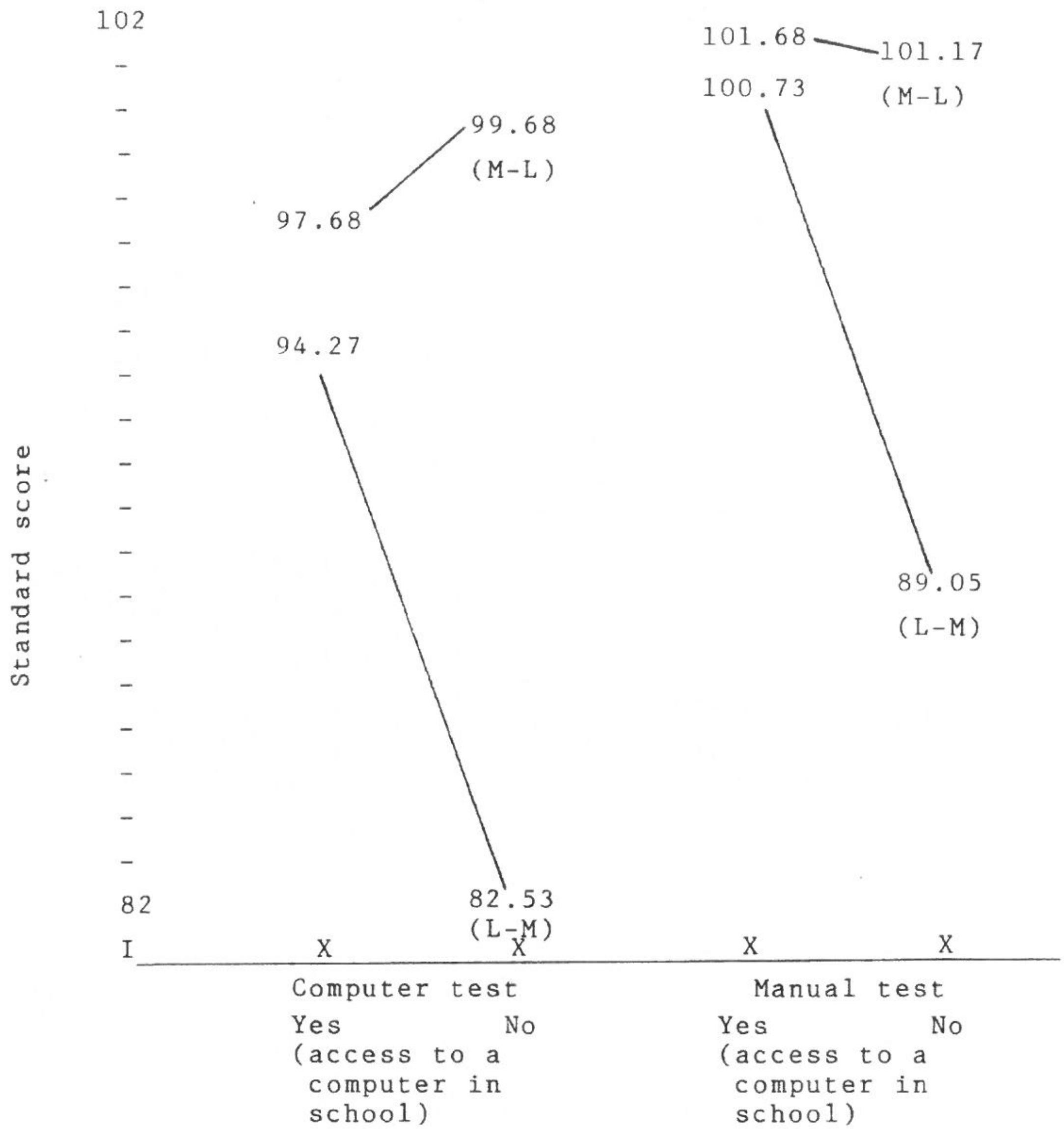

Figure 3. The influence of access to a computer in school, test sequence, and test format on test performance.

Table 5

Three-Way Analysis of Variance (Computer at Home x Test Sequence x Test Format) Summary Table

Source	SS	df	MS	F	p
Between subjects		97			
Experience	2,070.18	1	2,070.18	4.55	0.035
Test sequence	1,686.79	1	1,686.79	3.71	0.057
Experience x test sequence	21.30	1	21.30	0.04	
Error between	42,751.84	94	454.81		
Within subjects		98			
(FORMAT)	1,255.18	1	1,255.18	15.51	0.001
Experience x format	0.19	1	0.19	0.00	
Test x format	134.79	1	134.79	1.66	
Experience x test x format	38.00	1	38.00	0.46	
Error within	7,604.83	94	80.90		
Total		195			

with a computer at home scored higher than those who did not have a computer at home (see Figure 4). Although marginally significant, the test sequence influenced the scores. The manual-computer test sequence yielded higher scores than the computer-manual sequence under both the computer and manual test administrations ($\overline{X}$ = 98.17 and 90.28 for the computer score and $\overline{X}$ = 101.54 and 96.96 for the manual score).

As illustrated in Figure 4, those students who have access to a computer in their home did significantly better on both the manual- and computer-administered test than those students who did not have access to a computer in their home.

An additional finding was the influence of test sequence on test performance. Students who received the test manually before the computer-administered test (M-L) did better than those students who received the computer before the manual-administered test (L-M) (Figure 5).

Hypothesis 6--Analysis of the Influence of Gender Differences on Test Performance

The hypothesis that there would be a significant difference between male and female test performance under the different test administration conditions was tested by a three-way ANOVA with gender (male and female), test format (manual and computer), and test sequence (manual-computer and computer-manual) as the factors (see Table 6). The main effect of gender was not significant, $F(1, 94)$ =

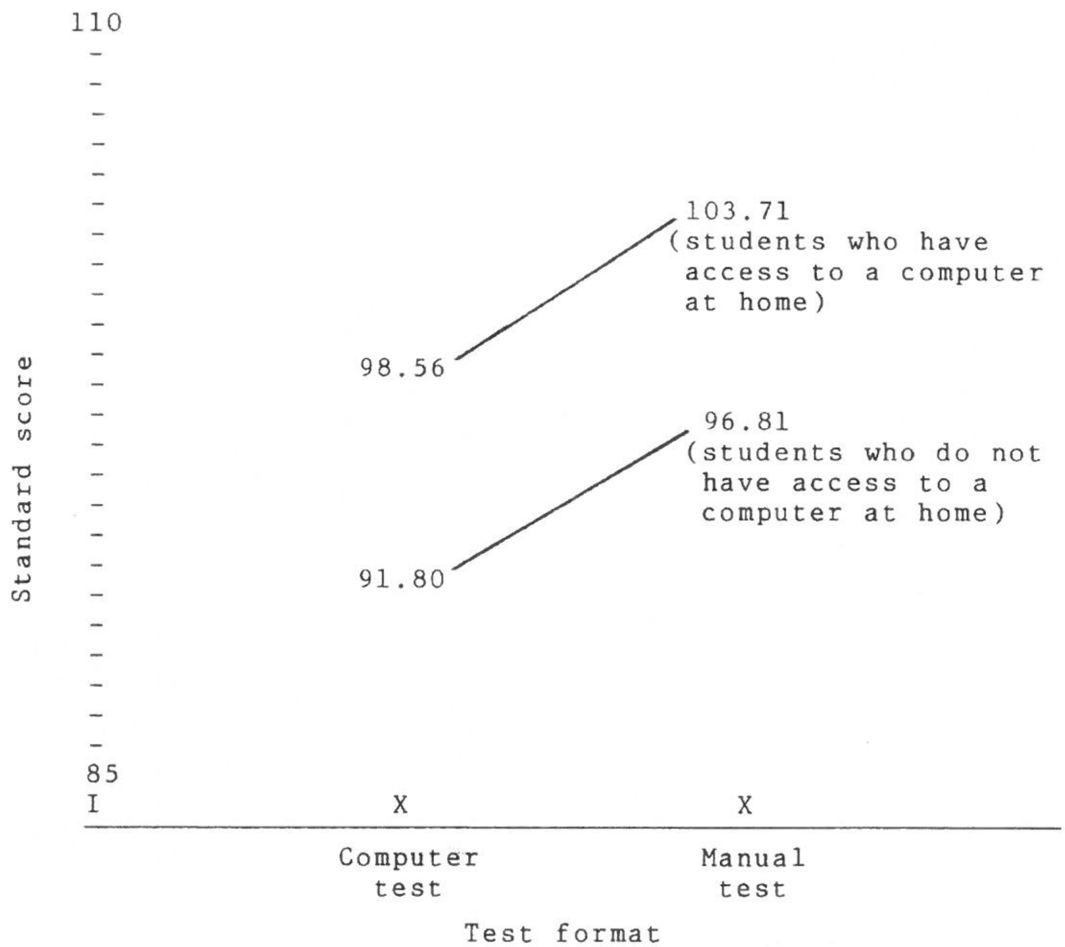

Figure 4. The influence of access to a computer at home and test format on test performance.

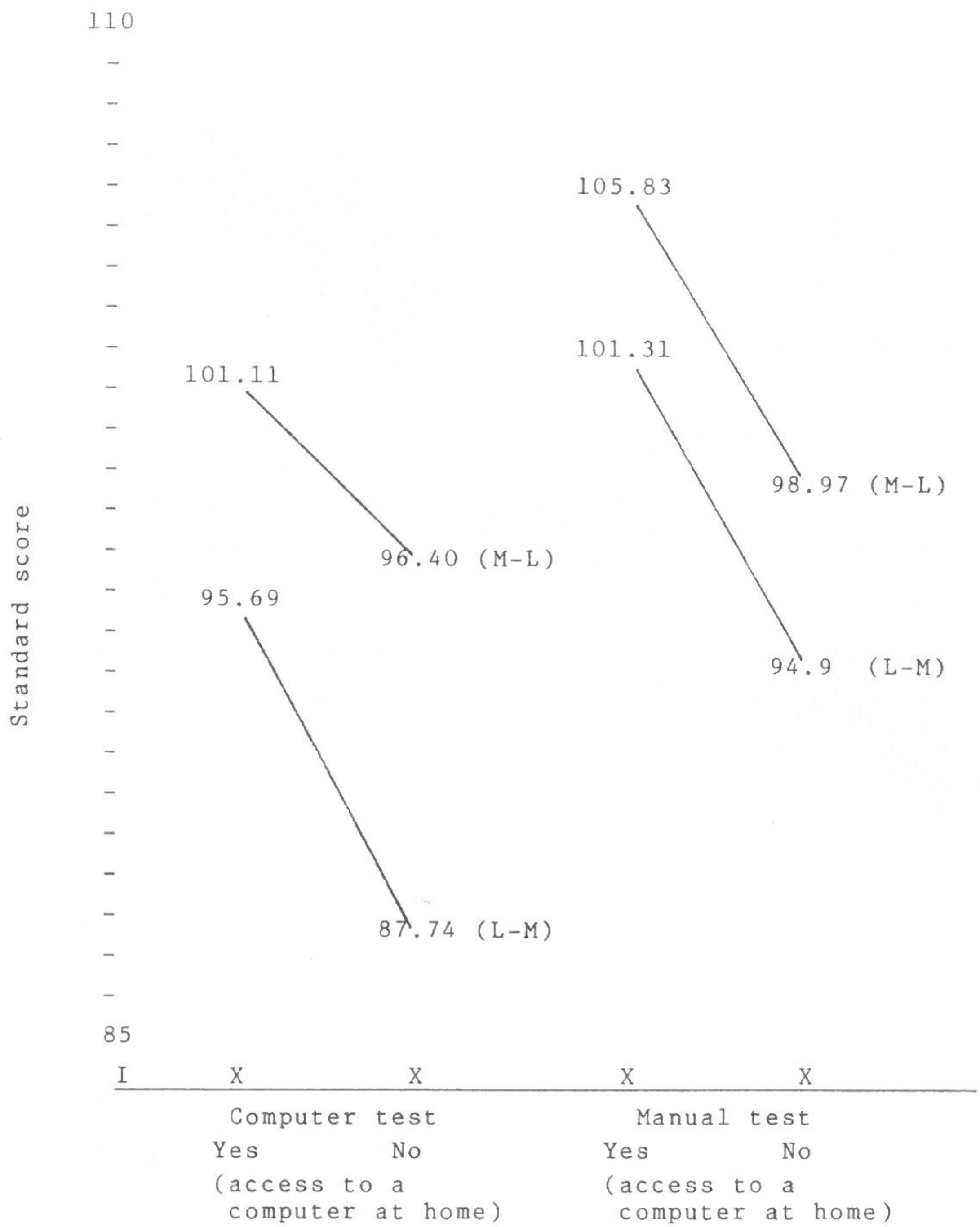

Figure 5. The influence of access to a computer in the home, test sequence, and test format on test performance.

Table 6

Three-Way Analysis of Variance (Gender x Test Sequence x Test Format) Summary Table

Source	SS	df	MS	F	p
Between subjects		97			
Gender	1.52	1	1.524	.003	.955
Test sequence	1,903.40	1	1,903.409	4.028	.047
Gender x sequence	208.71	1	208.713	.442	.507
Error between	44,416.45	94	472.515		
Within subjects		98			
Test format	1,255.18	1	1,255.184	15.528	.000
Gender x format	5.42	1	5.422	.067	.796
Test x format	133.66	1	133.662	1.653	.201
Gender x sequence x format	39.29	1	39.293	.486	.487
Error within	7,599.440	94	80.451		
Total		195			

.003. The main effect of the test sequence was significant, $F(1, 94) = 4.028$, $p < .05$. The main effect of the test format was statistically significant, $F(1, 94) = 15.526$, $p < .001$. The analysis did not yield a significant interaction effect.

The analysis presented in Table 6 indicated that male and female students scored the same, regardless of the test sequence or test format. Those students who were in the manual-computer test sequence performed better than those in the computer-manual sequence, regardless of gender and the test format. Those students who were in the manual-computer sequence did better, regardless of the test format or gender. The students performed better on the manual than the computer test, regardless of their test sequence or gender. Thus, Hypothesis 6 was rejected as gender did not influence test performance.

Chapter 5

DISCUSSION

This study addressed the variables of test reliability, age, gender, previous microcomputer experience, and preference for manual or computerized test administration on children's test performance under a manual and computer administration of the Peabody Picture Vocabulary Test-Revised.

Reliability Issues

This study found that there was no significant difference between the test-retest reliability coefficients derived from the computer and manual test administrations. However, when the alternate-form reliability coefficients derived in this study were compared to those reported in the Technical Supplement (Robertson & Eisenberg, 1981), a significant discrepancy occurred within the fifth-grade level L-M test sequence. The reliability found for the fifth-grade L-M test sequence was significantly lower than that reported in the Technical Supplement. One possible explanation for this significant discrepancy was the small number of subjects (n = 11) within the L-M group used in this study. It is possible that with such a small sample, unstable correlations enhanced a nonsignificant trend.

Unlike the Overton and Scott (1972) study, all the students in the experiment were able to reach a basal level. The concern over the student's ability to reach a basal had been previously overcome by Katz and Dalby (1981), who gave the subjects in their study a special training session on how to use the computer. The present study did not administer any special training sessions other than the sample items administered at the beginning of the PPVT-R. Given the performance of the students in this study and the cross-section of their abilities, it appears that the generation of students used in this study are well acquainted with the basic keyboard operations.

Influence of Test Sequence

This study found that the order of the test administration was significant. Specifically, those students who received the manual test before the computer test performed significantly better than those students who received the computer, followed by the manual test sequence. This trend was also reported by Knights et al. (1973). There were four factors which may have influenced the students' test performance under the computer-administered test. First, it was observed in this study that the computer format may have embellished the students' impulsive behaviors. Specifically, the students in this study were observed to respond to the test plates during the computer administration in an automatic manner. That is, when the students were presented with the pictures and the stimulus word, they responded immediately. Conversely, during the manual format, the students were observed to consider their choice before their response.

A factor which may have increased the students' tendency to respond quickly to the stimulus word during the computer administration was the 5-second time delay between stimulus words. The students may have increased their rate of responding under the computer administration as an attempt to decrease the time interval between stimulus words. The students' expectation that the computer would respond immediately to their keyboard commands probably stemmed from their experience with video games. However, given the excessive amount of graphics required for the PPVT-R program, it was not technically possible for it to "run" at the speed of a video game. An additional variable is that most educational and recreational software reward rapid responses and withhold reinforcement for delayed responses. The students may have been merely responding to the computer test administration in a manner which had been reinforced in the past.

The second factor which may have served to influence the discrepancy between test performance under the manual and computer conditions was the extent to which the students could change their item selection. Under the manual and computer conditions was the extent to which the students could change their item selection. Under the manual test condition, the subject could change the initial selection and choose a different picture. However, the computer test administration did not allow for the student to change his choice after the initial selection. Several students verbalized that they had made an error in their item selection and expressed a desire to change their response during both testing conditions. However, given the design of the computer system, it was not possible for the students to correct their first computer selection.

A third possible factor was that those students who received the computer test administration before the manual achieved a score which was more reflective of their true ability. That is, under the computer administration, the students were unable to react to the subtle clues which human test administrators reveal. Consequently, they were unable to increase their scores. Conversely, students who received the manual before the computer administration were "taught" the test, and had an advantage over students in the computer-manual test sequence. D. F. Johnson and Mihal (1973) also found that the subconscious biases and subtle cues of a test administrator can influence a student's test performance.

A fourth possible factor is that, although the students had been exposed to computers previous to this study, this was the first time they were administered an achievement test by computer, and thus, it presented a novel situation.

Influence of Previous Experience With Microcomputers on Test Performance

Analysis of the Previous Experience with Microcomputers Inventory found that 65.3% of the students reported having a computer system in their home. This study found that students who had computer systems in their home did significantly better on both the manual- and computer-administered tests than students who did not have computers in their homes. However, students with computers in their homes did significantly better on the manual test than the computer test. Thus, in this study students who had a computer in the home tended to exhibit greater academic performance. It appears that having a computer in the home is an academic advantage and not a technical advantage when taking an achievement test which is administered by a computer. Analysis of the Previous Experience with Microcomputers Inventory found that 6.1% of the students reported having a computer in their classroom, 28.5% reported there was occasionally a computer in their classroom, and 65.3% of the students reported that they did not have access to a computer in their school. Analysis of the test data indicated that those students with access to a computer system in their classroom did significantly better on both the computer and manual test administration formats. In addition, those students who had access to a computer system in their school did significantly better on the manual test than on the computer test. Hence, just as having exposure to a computer in the home is an academic advantage and not a technical advantage, students exposed to a computer in school appear to have an academic advantage and not a technical advantage.

Students' Test Administration Preference Questionnaire

The extent to which the students enjoyed the computer was evident on the Test Administration Preference Questionnaire. Specifically, 70.4% of the students expressed a preference for computer over human test administration. Whereas, 20.4% reported no preference for either human or computer test administrations for their next tests. Conversely, only 5% of the students expressed a desire to have their future tests administered by a human.

This study found that the students experienced no difficulty accepting a "talking computer." The students reported that they especially enjoyed the computer voice. In addition, many of the students reported that they liked the verbal praise that the computer offered during the test administration.

The extent to which the computer test became popular among the school children was further evident in that 20 children came to the library to request that they be allowed to take a test on the computer. The students could not be included in the study because they did not request to participate in the study until the 2nd week. It was hypothesized that the study gained in popularity when the students in the school realized that everyone in the study would get the chance to be tested by the talking computer.

Influence of Gender on Test Performance

Although Elwood and Clark (1978) found that female students performed significantly lower than male students under the electromechanical testing, this study found that the gender of the student did not influence the level of test performance under the computer or manual test administrations. One possible reason for the lack of a gender difference was that the students used in this study, both male and female, had received a level of exposure to microcomputers which eliminated the gender difference.

Limitations of the Study

Administration Time

The present study found that the tests administered by computer took longer than the tests which were administered manually. This trend was reported by Knights et al. (1973), who reported that the electromechanical administration took longer than the manual method of test administration. The principal reason for the extra length of the computer administration was due to the large amount of graphics which were displayed on the monitor. Since the design of this system, it is now possible to decrease the time required for the graphics display by two thirds with the use of an accelerator circuit board which can be installed in one of the expansion slots in the Apple IIe. Conversely, both this study and the Knights et al. (1973) study found that the scoring of the test was quicker when conducted by machine. It took approximately 30 seconds for the software program used in this study to score, interpret, and write a report. It took approximately 15 minutes to score the test manually.

Computer System

There were three limitations associated with the computer system utilized in this study. The first was the amount of heat generated by the computer, which was caused by the overheating of the hard disk drive. The system had to be turned off for approximately 1 hour each day so that it could cool down. The hard disk drives developed since the design of this computer system do not generate as much heat and can store three times as much data.

An additional concern involved the digital voice. Although the students reported that they enjoyed the voice, several of the words were difficult for them to understand. Many students found it necessary to have the computer repeat itself several times before they could understand some of the stimulus words.

The third limitation was that the students were unable to change their item selection once they had pushed a button on the response console. This was due to the design of the software program which did not provide a way in which the students could go back and change or correct their item choice. Future hardware systems are being designed to overcome the limitations of overheating of the hard disk drives and word recognition with digital voice.

Cost-Benefit Analysis

Although this study has shown that a PPVT-R administered by a computer system has acceptable test-retest reliability and is preferred over manual administration by 70.4% of the students, the question of the cost effectiveness of the use of such a system in the schools must be raised.

Research Costs

The total cost for developing the PPVT-R computer system used in this study was $12,800.00. This amount covered the cost of the hardware, the computer programmer, the consultants, the research data, the bata testing, the data collection, and miscellaneous expenses (see Appendix D).

Retail Cost

If the PPVT-R software program was to be marketed, it is estimated that the retail cost of the computer program would be $1,000 with a 15% discount for school districts which bought two or more systems. A school district which owned an Apple IIe system would have to purchase a 10-megabyte hard disk drive for approximately $700, a voice digital speech circuit board for $200, and a time clock circuit board for $100.00.

Employee Cost

To run the PPVT-R system requires an adult to turn the system on and to enter the student's name and year, month, and day of birth An adult is also required to turn the system off. Approximately 5 minutes are required to teach a computer-illiterate adult to manage the system. The system requires an adult to be in the same building with the student in case of a problem or emergency. The system may be managed by a school volunteer or teacher's aide with a cost of 0 to $5.00 per hour.

Analysis of Cost-Benefit

The following cost-benefit analysis is based on the following assumptions:

1. A school district could hire a school psychologist or speech pathologist for a total cost of $18,000 a year, including all of the employee benefits.

2. The individual administering the PPVT-R would be able to administer a PPVT-R every 20 minutes and score the protocols during the 5-minute breaks between tests. The tester could administer 9,000 PPVT-Rs a year, assuming no

sick days off, a constant test administration rate, and no mistakes.

3. The school district owned an Apple IIe system.

4. The school district bought the hardware and software at the prices listed above for a total cost of $2,000, and then had an adult in the building for a maximum cost of $4,500 per year, assuming that the adult works 5 hours a day at a rate of $5 an hour for 180 days.

5. The computer system ran 8 hours a day and administered, scored, and interpreted 5,400 tests a year. This analysis is not based on the use of an accelerator card.

Given that the school district found a school psychologist or speech pathologist for $18,000 per year who could administer 9,000 tests per year and the computer system cost $3,900 and administered 5,400 tests a year, the district would save $1.27 for each computer test administration. The system would pay for itself in 6 months. Furthermore, the computer system could be used for other activities in addition to administering tests.

Hence, it can be seen that, not only is it acceptable and in many cases preferable for a PPVT-R to be administered by computer, but that the benefit of future testing systems in terms of cost and greater reliability will significantly outweigh the current practice of having professionals spend valuable time screening students for possible academic delays.

Future Research

Future research should address the variables which influence test sequence x test format effect. It is suggested that studies which investigate this phenomenon use a large sample of students with a variety of academic abilities and with a variety of test response formats (i.e., pressing button, pointing, verbal responses).

Another study could investigate the effects of a decrease in the time span between stimulus words on test performance.

Future research is needed to examine the extent to which socioeconomic status may be influencing a student's achievement instead of access to a computer system in the home. The type of home computer system and home software programs should be examined as well as the amount of time the student spends with the system. The finding of this study that students exposed to computer systems in either home or school achieved a higher level of academic performance on the PPVT-R than students who did not have the exposure may discredit the socioeconomic status variable as a significant confounding influence. However, additional research is needed in this area.

Another area for further research is based on the evidence that students who have access to computers in their classroom score significantly higher on achievement tests than children who do not. Perhaps those teachers who do not have computers in their classrooms are presenting the curriculum in a format which does not enhance test performance. Moreover, teachers who actively pursue the use of computers in their classrooms may teach the curriculum in a manner which enhances test performance.

A third possibility is that there is an unequal distribution of computers in school systems. Hence, further research should examine the computer distribution and student access time in schools. For example, which grade level and type of classrooms (i.e., gifted, special education, low achievers) are most likely to have computer systems, and is there a differential increase in academic performance after exposure to computer systems which run educational software based on the type of students (i.e., learning disabled versus gifted versus regular education).

An investigation could be conducted which addresses the influence that a computer system itself appears to have on increasing academic performance. Such a study would be similar to the studies researched by Kulik et al. (1985). It would consist of dividing a group of low academic achievers into two groups. One group would receive the typical drill and practice of educational concepts offered on most educational software. The other group would not be given any exposure to the computer. After a period of time, the two groups would be compared on test performance under both a manual and computer achievement test.

The extent to which test preference (computer versus manual) can influence test performance on manual- and computer-administered tests could also be researched.

Given the findings of this study, it appears that computer administration of achievement tests offers the potential of saving school districts a significant amount of money. In short, this study supports the contention that a computer can administer tests to children in a structured environment with minimal adult supervision. The computer system designed for this study needed adult supervision because the system had to be turned on and the child's name and date of birth entered. Future computer systems could be designed to allow for the child to enter the data, or an adult could enter the data into a file system which would store the information for future use. Therefore, the advantage of the computer over human test administration in terms of cost effectiveness is substantial, given that once the program is developed, no human interaction is required for the administration, scoring, interpretation, or reporting of the test results.

The cost-benefit analysis conducted for the computer system developed for this study indicated that the benefits of computer administration far outweigh the costs of manual administration. It is important to bear in mind that this study used a test which was designed to be administered manually. Conversely, future computerized test administration systems will be based on tests which are designed and normed for computer administration. Hence, the initial research and development costs are not known at this time. However, given that the structure of the hardware system and software programs have been developed in this study, the initial capital investment for future computerized test administration systems will be greatly reduced.

An additional feature of computer-based test administrations is that they could be curriculum embedded. For example, the computer system could be programmed to not only administer a test of mathematics skill but assess the concepts (e.g., carrying numbers in division problems) for individuals who score below a set point. Hence, the computer-administered test could not only assess the level of academic performance but knowledge of concepts. The benefit of such a testing system is that all children, not just those referred to the school psychologist, could receive an assessment of their level of academic skill and knowledge of academic concepts. The testing system could easily serve as a standardized screening instrument with those children who perform below grade level being referred to the school psychologist for additional evaluation. Understandably, such computer-based testing systems would require a new field of test validation. Specifically, such systems, given the nature of the interactive assessment process, would require that norms be developed for the computerized test administrations just as norms have been developed for the manual versions in the past.

Conclusions

Given that a 64K Apple IIe computer with the addition of a 10-megabyte hard disk drive, a voice digitizer, and a four-button response console were all that was needed to run the PPVT-R, the cost of such a system is well within the reach of school districts. However, the true benefit of systems such as this will not lie in the administration of the PPVT-R, but in achievement and personality tests which can be specifically tailored to use the capabilities of microcomputers. An additional benefit inherent in a computer system such as the one used in this study is that the system can administer the test without the presence of a psychologist, speech and language clinician, or classroom teacher.

There is little question that as computer systems are embedded into the teaching curriculum the possibility of a computer administering testing system being accepted by the educational community is increased. Within the near future, such testing systems will most probably appear in the educational arena. This writer can perceive of a time when the national achievement tests may be administered by computer systems in the schools.

Moreover, given the results of this study, it appears that it is now feasible for a computer program to be developed which would reliably and economically administer the achievement tests frequently used in education. In addition, given that the students who were exposed to computers at home and in their classrooms consistently scored higher on the manual rather than the computer testing sytem, such computer testing systems do not appear to be a handicapping condition for those children who did not have access to computerized teaching systems.

REFERENCES

REFERENCES

American Psychological Association. (1986, February). Monitor, 17(2), 74-78.

Angle, H. V., & Carroll, J. (1979). A computer interview language: Programming the on-line interactive computer. Behavior Research Methods and Instrumentation, 11, 379-383.

Angoff, W. H. (1953). Test reliability and effective test length. Psychometrika, 18, 1-14.

Apple peripherals. (1984). Carrollton, TX: Applied Engineering.

Beaumont, J. G. (1985). The validity of the category test administered by on-line computer. Journal of Clinical Psychology, 31, 485-462.

Birtles, C. J., Sambrooks, J. E. MacCulloch, M. J., & Holland, P. (1972). An inexpensive computer assisted psychometric system. Medical and Biological Engineering and Computing, 10, 145-152.

Byers, A. P. (1981). Psychological evaluation by means of an on-line computer. Behavior Research Methods and Instrumentation, 13,585-587.

Calvert, E. J., & Waterfall, R. C. (1982). A comparison of conventional and automated administration of Raven's standard progressive matrices. International Journal of Man-Machine Studies, 17, 305-310.

Carr, A. C., & Ghosh, A. (1983). Accuracy of behavioral assessment. British Journal of Psychiatry, 142, 66-70.

Davidheiser, R. B. (1982). The domain voice analysis system for speech pathology (Report No. 0195-4210). Lebanon, PA: Data System Company.

Dunn, L.,& Dunn, L. (1981). Peabody picture vocabulary test-revsied manual: Forms L and M. Circle Pines, MN: American Guidance Service.

Dunn, T. G., Lushene, R. E., & O'Neil, H. F. (1972). Complete automation of the MMPI and a study of its response latencies. Journal of Consulting and Counseling Psychology, 39, 381-387.

Elwood, D. L. (1969). Automation of psychological testing. American Psychologist, 24, 287-289.

Elwood, D. L. (1972a). Automation versus face-to-face intelligence testing: Comparison of test-retest reliabilities. International Journal of Man-Machine Studies, 4, 363-369.

Elwood, D. L. (1972b). Automated WAIS testing correlated with face-to-face WAIS testing: A validation study. International Journal of Man-Machine Studies, 4, 129-137.

Elwood, D. L. (1972c). Test-retest reliability and cost analyses of automated and face-to-face intelligence testing. International Journal of Man-Machine Studies, 4, 1-22.

Elwood, D. L. (1973). Reliability of automated intelligence testing using a three-month test-retest interval. Intrnational Review of Applied Psychology, 22, 157-162.

Elwood, D. L., & Clark, C. L. (1978). Computer technology: Computer administration of the Peabody picture vocabulary test to young children. Behavior Research Methods and Instrumentation, 10, 43-46.

Elwood, D. L., & Griffin, H. R. (1972). Individual intelligence testing without the examiner: Reliability of an automated method. Journal of Consulting and Clinical Psychology, 38, 9-14.

Finney, J. C. (1966). Programmed interpretation of MMPI and CPI. Archives of General Psychiatry, 15, 75-81.

Fowler, R. D. (1985). Landmarks in computer-assisted psychological assessment. Journal of Consulting and Clinical Psychology, 53, 48-57.

Geyde, J. L., & Miller, E. (1970). Developments in automated testing systems. In P. Mittler (Ed.), The psychological assessment of mental and physical handicaps (pp. 735-760). London: Methuen.

Gilberstadt, H., Lushene, R. E., & Buegel, B. (1976). Automated assessment of intelligence: The TAPAC test battery and computerized report writing. Perceptual and Motor Skills, 43, 627-635.

Gynther, M. D., Altman, A., & Sletten, I. W. (1973). Replicated correlates of MMPI two-point code types: The Missouri actuarial system. Journal of Clinical Psychology, 29, 263-289.

Hedlund, J. L., Vieweg, B. W., & Cho, D. W. (1985a). Mental health computing in the 1980s: I. General information systems and clinical documentation. Computers in Human Services, 1(1), 3-33.

Hedlund, J. L., Vieweg, B. W., & Cho, D. W. (1985b). Mental health computing in the 1980s: II. Clinical applications. Computers in Human Services, 1(2), 5-25.

Johnson, D. F., & Mihal, W. L. (1973). Performance of blacks and whites in computerized versus manual testing environments. American Psychologist, 28, 694-699.

Johnson, J. H., & Williams, T. A. (1980). Using on-line computer technology to improve service response and decision-making effectiveness in a mental health admitting system. In J. B. Sindowski, J. H. Johnson, & T. A. Williams (Eds.), Technology in mental health care delivery systems (pp. 5020-5031). Norwood, NJ: Ablex.

Katz, L., & Dalby, J. T. (1981). Computer-assisted and traditional psychological assessment of elementary-school-aged children. Contemporary Educational Psychology, 6, 314-322.

Klett, C. J., & Pumroy, D. K. (1971). Automated procedures in psychological assessment. In P. McReynolds (Ed.), Advances in psychological assessment (pp. 14-39). Palo Alto, CA: Science & Behavior Books.

Klinge, V., & Rodziewicz, T. (1976). Automated and manual intelligence testing of the Peabody picture vocabulary test on a psychiatric adolescent population. International Journal of Man-Machine Studies, 8, 243-246.

Klingler, D. E., Johnson, J. H., & Williams, T. A. (1976). Strategies in the evaluation of an on-line computer-assisted unit for intake assessment of mental health patients. Behavior Research Methods and Instrumentation, 8, 95-100.

Knights, R. M., Richardson, P. H., & McNarry, L. R. (1973). Automated versus clinical administration of the Peabody picture vocabulary test and the coloured progressive matrices. American Journal of Mental Deficiency, 28, 223-225.

Kulik, J. A., Kulik, C. C., & Bangert-Drowns, R. L. (1985). Effectiveness of computer-based education in elementary schools. Computers in Human Behavior, 1, 59-74.

Lushene, R. E. (1981). Development of a psychological assessment system. In H. J. Hefferman (Ed.), Proceedings of the fifth annual symposium on computer applications in medical care (pp. 235-250). New York: Institute of Electrical and Electronics Engineers.

Magellan light pen system II manual. (1985). Indianapolis, IN: Magellan Computer, Inc.

McCall, R. B. (1980). Fundamental statistics for psychology. New York: Harcourt, Brace, Jovanovich.

Moreland, K. L. (1985). Computer-assisted psychological assessment in 1985: A practical guide. Minneapolis, MN: National Computer Service, Professional Assessment Services.

Overton, G. W., & Scott, K. G. (1972). Automated and manual intelligence testing: Data on parallel forms of the Peabody picture vocabulary test. American Journal of Mental Deficiency, 76, 639-643.

Paitich, D. (1973). A comprehensive automated psychological examination and report (CAPER). Behavioral Science, 18, 131-136.

Robertson, G. J., & Eisenberg, J. L. (1981). Technical supplement--Forms L and M: Peabody picture vocabulary test-revised. Circle Pines, MN: American Guidance Service.

Schwartz, M. D. (Ed.). (1984). Using computers in clinical practice: Psychotherapy and mental health applications. New York: Haworth Press.

White, D. M. (1983). An assessment of the comparability of a computer administration and standard administration of the mmpi. Unpublished master's thesis, University of Alabama.

APPENDICES

APPENDIX A

INTRODUCTORY LETTER, PERMISSION SLIP, AND SUBJECT CONSENT FORM

CSPP
CALIFORNIA SCHOOL OF PROFESSIONAL PSYCHOLOGY

BERKELEY
FRESNO
LOS ANGELES
SAN DIEGO
PRESIDENT'S OFFICE

FRESNO CAMPUS
1350 M ST
FRESNO CA 93721
(209) 486 8420

My name is Terrance Lichtenwald. I am conducting educational research under the supervision of the California School of Professional Psychology in Fresno. My study will investigate the influences that a test administered by a computer has on a child's test performance. The students selected will take the Peabody Picture Vocabulary Test by computer and by the traditional method. The Peabody Picture Vocabulary Test measures a child's knowledge of words.

I will also ask each child to fill out two questionnaires. Before a child is tested, he or she will be asked questions about how much experience he or she has had with computers. After a child has taken the test by both the computer and traditional methods, he or she will be asked to respond to a brief questionnaire describing his or her feelings about each testing method.

As a parent, you are well aware of the extent to which computers are used in daily life. The importance of knowing the extent to which a child's experience with computers influences his or her ability in school is of major concern to educators today. The findings from my research will be of benefit to the people in the schools who test and teach children. The identities of the children and their individual test scores will be kept in strict confidence. Only the summarized results of this study will be made known to the schools. Be assured that your child's identity will remain anonymous.

This study has been approved as one which does not infringe on you or your child's rights as a participant or endanger you or your child in any way. As a participant in the study, your child will gain the experience of interacting with an Apple Computer. In addition, information regarding your child's performance will be made available to you per your request by contacting me at the address listed above or through the principal of your child's school.

I would sincerely appreciate your permission to place your child's name on the list from which a group of children will be randomly drawn to participate in my study. Again, your child's identity will be protected and only the summarized findings will be made known.

If you wish to have your child's name placed on the list of possible participants in this study, please complete the enclosed Consent Form for Research Subjects and have your child return it to his or her classroom teacher. Thank you for your cooperation.

Sincerely,

/s/ Terrance Lichtenwald

Terrance Lichtenwald, MS, MA

CSPP
CALIFORNIA SCHOOL OF PROFESSIONAL PSYCHOLOGY

BERKELEY
FRESNO
LOS ANGELES
SAN DIEGO
PRESIDENT'S OFFICE

FRESNO CAMPUS
1350 M ST
FRESNO CA 93721
(209) 486-8420

Your child has been selected to participate in a study which will investigate the influences that a test administered by computer has on a child's test performance. Your child will take the Peabody Picture Vocabulary Test by computer and by the traditional method. The Peabody Picture Vocabulary Test measures a child's knowledge of words. Your child will also be asked to fill out a Previous Experience With Computers Inventory and a Test Administration Preference Questionnaire.

The findings from my research will be of benefit to the people in the schools who test and teach children. Your child's identity and his or her indiviidual test scores will be kept in strict confidence. Only the summarized results of this study will be made known to the schools.

This study has been approved as one which does not infringe on you or your child's rights as a participant or endanger you or your child in any way. As a participant in this study, your child will gain the experience of interacting with an Apple computer. In addition, information regarding your child's performance will be made available to you upon your request by contacting me at the address listed above or through the principal of your child's school.

I would sincerely appreciate your permission to allow your child to participate in this study. If you do not wish to have your child's name placed on the list of participants in this study, please complete the following form and have your child return it to his or her classroom teacher by tomorrow (Wednesday, May 28) morning. Thank you for your cooperation.

Sincerely,

/s/ Terrance G. Lichtenwald

Terrance G. Lichtenwald, MA, MS
CSPP-Fresno

I do not wish to have my child participate in this study.

____

CSPP
CALIFORNIA
SCHOOL OF
PROFESSIONAL
PSYCHOLOGY

BERKELEY
FRESNO
LOS ANGELES
SAN DIEGO
PRESIDENT'S OFFICE

FRESNO CAMPUS
1350 M ST
FRESNO, CA 93721
(209) 486-8420

CONSENT FORM FOR RESEARCH SUBJECT

I, ______________________________, give my consent to Terrance Gordon Lichtenwald, who is listed below as the principal investigator, to collect psychological data on my child ______________________ for this research project. I am aware that the data collected will be kept confidential and that the principal investigator will follow the American Psychological Association (APA) Ethical Standards, including those for research with human subjects.

____ Date
Signature of Legal Guardian

/s/ Terrance G. Lichtenwald 5/12/86
Signature of Principal Investigator Date

Name of the Research project: AN INVESTIGATION OF THE VALIDITY, RELIABILITY, AND ACCEPTANCE BY CHILDREN OF A MICROCOMPUTER ADMINISTRATION OF THE PEABODY PICTURE VOCABULARY TEST-REVISED

Questions and concerns about the conduct of this research may be addressed to the California School of Professional Psychology-Fresno (CSPP-F) Research Committee at 1350 M Street, Fresno, CA 93721, 486-8420.

AT THE CONCLUSION OF THE RESEARCH PROJECT, THE PRINCIPAL INVESTIGATOR FILES THESE FORMS (OR AN ADAPTATION OF THIS FORM) WITH THE CSPP-F RESEARCH COMMITTEE.

____
Form 5 Date filed

APPENDIX B

PREVIOUS EXPERIENCE WITH MICROCOMPUTERS

INVENTORY

Previous Experience With Microcomputers

Inventory

Name:____________________________________ Boy___ Girl___

Directions:

Please circle the best answer that tells how you feel.

Questions:

1. Do you have a computer at home?

   A. Yes
   B. No

2. If you have a computer at home, is it

   A. an Apple
   B. an Atari
   C. a Commadore
   D. a Texas Instruments
   E. an Adam
   F. an Epson
   G. an I.B.M.
   H. a Peach
   I. a Rainbow
   J. a Zenith
   K. a Macintosh
   L. other____

3. Do you play with the computer at home?

   A. Yes
   B. No

4. How much do you play with the computer?

   A. Every day
   B. 5-6 days a week
   C. 3-4 days a week
   D. 1-2 days a week
   E. Never

5. On a scale/line from 0 to 10 with 0 being never and 10 being all the time, how often do you play with a computer?

   0____________________________5____________________________10
   Never All the time

6. How long have you had a computer at home?

   A. All my life
   B. 4 years
   C. 3 years
   D. 2 years
   E. 1 year
   F. Never

7. Do you like your computer at home?

   A. Yes
   B. No
   C. I do not have a computer at home.

8. Do you have a computer in your school?

   A. Yes
   B. No

9. Is there a computer in your classroom?

   A. Yes
   B. No
   C. Sometimes

10. Do you get to work with the computer in your school?

    A. Yes
    B. No
    C. Sometimes
    D. I do not have a computer in my school.

11. How often do you get to work with the computer in your school?

    A. Every day
    B. 5 to 6 days a week
    C. 3 to 4 days a week
    D. 1 to 2 days a week
    E. Never
    F. I do not have a computer in my school.

12. Do you like to work with the computer at school?

    A. Yes
    B. No
    C. Sometimes
    D. I do not have a computer in my school.

13. When you work with the computer at school do you share the computer with other students?

A. Yes
B. No
C. Sometimes
D. I do not have a computer in my school.

14. Have you ever worked with the computer in your school alone?

A. Yes
B. No
C. I do not have a computer in my school.

15. Do you like having computers in your school?

A. Yes
B. No
C. I do not have a computer in my school.
D. Please tell why you said yes or no__________________
____
____

16. If you do not have a computer in your school, would you like to have one?

A. Yes
B. No
C. I have a computer in my school.
D. Please tell why you said yes or no__________________
____
____

17. Have you ever gone to a game room to play video games?

A. Yes
B. No

18. Do you like to play video games?

A. Yes
B. No

18. When you play video games, do you usually play be yourself?

A. Yes
B. No
C. I don't play video games.

19. Do you like to go to the video games room with your friends?

    A. Yes
    B. No

THANK YOU

APPENDIX C

TEST ADMINISTRATION PREFERENCE QUESTIONNAIRE

Test Administration Preference

Questionnaire

Name______________________________________ Boy___ Girl___

Directions:

Please circle the best answer that tells how you feel.

Questions:

1. Which test took longer?

   A. The computer test
   B. The test given to me by the person
   C. No difference

2. Which test was harder?

   A. The computer test
   B. The test given to me by the person
   C. No difference

3. Which of the tests was most like play?

   A. The computer test
   B. The test given to me by the person
   C. No difference

4. Which of the tests was most like work?

   A. The computer test
   B. The test given to me by the person
   C. No difference

5. Which of the tests do you think you did the best on?

   A. The computer test
   B. The test given to me by the person
   C. No difference

6. Was the computer test like any of the computer games you have played before?

   A. Yes
   B. No

7. Which way would you like to take your tests in the future?

   A. By computer
   B. Have someone give the test to you
   C. Don't care if a computer or person gave you the test.

8. Do you think the computer test was like the programs/ games you use at school?

   A. Yes
   B. No
   C. I don't use a computer at school.

9. Do you think the computer test was like the programs/ games you use at home?

   A. Yes
   B. No
   C. I don't have a computer at home.

10. Did you like the woman who gave you the test?

   A. Yes
   B. No
   C. I don't know.

11. Did you like the woman in the computer who gave you the test?

   A. Yes
   B. No
   C. I don't know.

12. Did you like the woman in the room with the computer?

   A. Yes
   B. No
   C. I don't know.

13. Using the computer was

   A. Fun
   B. Boring
   C. Not sure

14. The words on the computer were

   A. Easy to understand
   B. Hard to understand
   C. Not sure

15. Have you ever taken a test by computer before?

    A. Yes
    B. No

16. Tests given by a computer are

    A. Harder than tests given by a person
    B. Easier than tests given by a person
    C. Can't decide

17. While taking the test on the computer I

    A. Did not care if I did well
    B. Tried hard to do my best
    C. Can't decide

18. While taking the test given by the person I

    A. Did not care if I did well
    B. Tried hard to do my best
    C. Can't decide

19. Did you like the computer which gave you the test?

    A. Yes
    B. No
    C. I don't know.
    D. Please tell why you said yes, no, or I don't know.

____
____
____

THANK YOU

APPENDIX D

RESEARCH COSTS

RESEARCH COSTS

	Item	Cost
A.	Hardware Computer hardware including the Apple IIe system and the video camera system, light pen, and circuit boards used in development	$3,000.00
B.	Employee Salaries The salary for the computer programmer	3,500.00
C.	Research Review of the hardware and software designs, development of speech system and double high resolution graphics, and review of literature	2,500.00
D.	Consultants Electrical engineering, system analyst, speech pathologists, school psychologists	1,300.00
E.	Bata Testing Rewriting a section of the speech program, shipping of the computer hardware	1,000.00
F.	Data Collection Test administrators, gift to school site	500.00
G.	Miscellaneous Transportation, telephone bills, shipping of computer components to consultants	1,000.00
	Total	$12,800.00

APPENDIX E

PERMISSION TO USE THE PEABODY PICTURE VOCABULARY TEST-REVISED FORMS L AND M

AGS®

American Guidance Service, Inc.
John P. Yackel, President

September 15, 1986

Terrance G Lichtenwald
509 Baldwin Avenue Apt C3
Waukegan Illinois 60085

Dear Mr. Lichtenwald

The authors of the Peabody Picture Vocabulary Test-R (PPVT-R) are out of the country. As publishers of this test we have been authorized to act in their behalf regarding permission requests.

We hereby grant you permission to use the PPVT-R in your doctoral dissertation. Should you wish to reproduce actual test copy in your dissertation, it would be necessary to again contact us.

I am certain that both the authors and our Test Division will be interested in reviewing your dissertation findings.

Sincerely

Dorothy J Morstad

Dorothy J Morstad
Assistant to the President

DJM/jam

Partners in developing human potential

AGS® / Publishers' Building • Circle Pines, Minnesota 55014-1796 • Telephone: (612) 786-4343