12-1989

Comparison of Traditional and Computer-Assisted Administrations of the PPVT-R with Trainably Mentally Impaired Students

Wanda Balla

COMPARISON OF TRADITIONAL AND COMPUTER-ASSISTED ADMINISTRATIONS OF THE PPVT-R WITH TRAINABLY MENTALLY IMPAIRED STUDENTS

by

Wanda Balla

A Project Report
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Specialist in Education
Department of Psychology

Western Michigan University
Kalamazoo, Michigan
December 1989

COMPARISON OF TRADITIONAL AND COMPUTER-ASSISTED ADMINISTRATIONS OF THE PPVT-R WITH TRAINABLY MENTALLY IMPAIRED STUDENTS

Wanda Balla, Ed.S.

Western Michigan University, 1989

A group of 20 trainably mentally impaired students, ages 6-24, was randomly assigned to two groups, one of which received the Peabody Picture Vocabulary Test-Revised (Dunn & Dunn, 1981), manual administration, and the other a computer-assisted version, which required a switch closing response. Two weeks later, the order of administration was reversed. The results showed a difference of -1.73 points in the manual to computer administration, and a -3 point difference from computer to manual administration. Overall difference between both administrations was -2.36 points, favoring the manual administration. Recommendations include systematic evaluation and/or training of TMI or lower-functioning students before using a method of evaluation requiring this type of response.

ACKNOWLEDGEMENTS

To my committee members, Dr. Howard Farris, Chairman, Dr. Galen Alessi and Dr. Steve Ragotzy, thanks for your encouragement, criticism, questions and prompts. You are all truly facilitators.

Thanks to Jim Yanna who helped with the computer program and provided much feedback, advice and support.

And finally, thanks to Nick Balla and Mike Murphy whose tolerance and help were invaluable in completing this project.

Wanda Balla

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

Education and psychological evaluation of severely impaired children is a challenging and often controversial task. It is not uncommon to find a variety of approaches being used to accomplish these evaluations with those involved claiming that their approach is best for determining the needs of these persons. Fortunately, with the passage of laws dealing with the education of the handicapped, like Public Law 94-142 (Davis, 1986) demands for valid and comprehensive evaluations have not only increased but documentation is expected and often required. It is expected that the assessment process will produce data regarding a variety of needs of these youngsters including academic, vocational, and often social. Part of the problem of addressing assessment needs of the severely impaired is preparation and training of professional personnel. For example, Matey (1985) has suggested: "School psychologist training programs do not deal extensively, if at all, with issues in assessment and programming for children functioning at preschool developmental levels. Two additional areas of deficient training are experience in interacting with Severely and Profoundly Handicapped (SPH) children in developing teaching strategies for addressing special learning problems in educational settings" (p. 285).

A second and substantial problem is the type of information produced from the typical or traditional assessment process.

Proponents of this (traditional) approach usually attempt to describe the child in terms of IQ and/or mental age. This results in two problems: (1) traditional tests are usually far beyond the abilities of this group and (2) the testing results in little discussion of teaching strategies or instructional objectives for the person being tested. The importance of linking classroom instruction and assessment is discussed by Bagnato and Neisworth (1981). Both the developmental approach, as discussed by Bagnato and Neisworth and the functional approach, as discussed by Brown et al. (1985) have been deemed appropriate for use with this group (Matey, 1985). The developmental approach is characterized by construction and validation based on normative standards, and proposes (a) behavior changes which follow a heirarchy that is fairly fixed and (b) behavior acquisition that moves from simple to complex; more complex behavior is the result of modification or coordination of basic component behaviors (Barnett, 1983).

The functional approach proposes to use the "criterion of ultimate functioning" for evaluating severely handicapped students as stated by Barnett (1983):

> The criterion of ultimate functioning refers to the ever changing, expanding, localizing, and personalizing cluster of factors that each person must possess in order to function as productively and independently as possible in socially, vocationally, and domestically integrated adult community environments. (p. 164)

The issue of whether to use standardized tests at all has been debated and criticized (Cruickshank, 1986), as well as lauded. Sattler (1988) states:

> Intelligence testing is part of the science of human abilities. The science is young and developing. Although there are weaknesses in the technology, there is a solid scientific basis for the practice of mental ability testing [Carroll & Horn, 1981]. Furthermore, the differential psychology of cognitive abilities can do much to improve the human condition. (pp. 79-80)

The use of standardized tests provides important evidence to support inferences about cognitive functioning (Barnett, 1983). Sattler (1974) mentions some problems that may exist when testing handicapped children:

> First, communication difficulties may exist. Speech and hearing deficiencies may produce false impressions about the child's intellectual ability. Second, the child may become fatigued easily because he is unaccustomed to concentrated work for long periods of time. Third, if there are attention difficulties, they may be associated with physical deficiencies. Fourth, rapport difficulties may occur with those children who have heightened dependency. (p. 80)

Before choosing the mode of assessment which will best suit the client's needs a number of things should occur. First, it is important to evaluate the child's sensory-motor capabilities which could be done by asking teachers and/or parents about signs, symptoms or gestures that the child uses and the meaning of these activities. Secondly, information about a child's vision, hearing, physical condition and health is critical. The following areas are specifically suggested by Sattler (1974): "a) vision, hearing, speech, sitting balance, arm-hand use, b) reading and writing skills (for a school-aged child), and c) ability to indicate yes or no by either verbal or non-verbal means" (p. 97).

When the choice is made to use psychometric tests as part of the evaluation, the examiner should examine the instruments for

developmental information to assist with placement and/or determination of general skill levels. There is a critical need to obtain descriptive data and to avoid only using a numerical summary of the child's level of functioning. Basic considerations should include "A knowledge of evaluation procedures and instruments which address the sensory and/or response limitations of the SPH group, the ability to assess children's functional skills and to generate instructional objectives and educational strategies" (Matey, 1985, p. 285).

The assessment of one such "functional skill," i.e., receptive language, important for communication and language development for the severely impaired may be measured with the Peabody Picture Vocabulary Test-Revised (PPVT-R) (Dunn & Dunn, 1981). Discussion of the PPVT-R in terms of its validity, appropriateness and usefulness are presented below (American Psychological Association [APA], 1985).

The PPVT-R is a revised edition of the PPVT (Dunn, 1959) and is a norm-referenced, wide range, power test of hearing vocabulary. It is available in two parallel forms, each form containing five training items, followed by 175 test items arranged in order of increasing difficulty (Dunn & Dunn, 1981). The test is designed for persons 2 1/2 through 40 years old, who can see and hear reasonably well and understand standard English. Standardization was conducted on a national basis with a sample of 4200 children and adolescents and 800 adults. Raw scores are converted to age referenced norms, standard scores, and percentile equivalents.

Prout and Schwartz (1984) found that the tendency of the PPVT-R was to yield "too low" estimates, compared to the WAIS and WAIS-R scores. Therefore, they recommend that the test should not be used as a sole measure of intellectual functioning. Also, the PPVT-R is most appropriate as a measure of verbal comprehension, which does not require verbal expression, that may be used in conjunction with other tests in a psychoeducational battery (Naglieri, 1982).

Other advantages of PPVT-R are that subjects are not required to read and write or to speak. Also, because neither pointing nor oral responses are essential, the test can be used with even severely motor-handicapped persons such as those with cerebral palsy. Furthermore, it has been recommended that the alternate forms of the PPVT-R could be used with reasonable confidence in estimating the vocabulary competencies of adults with developmental disabilities (Groeneweg, Conway, & Stan, 1986). McCallum (1985) found that "comparability of the alternate forms seems well established" (pp. 1126-1127), as did Worthing (1984).

In an attempt to increase the validity of a standardized test with a population of moderately to severely handicapped students, adaptations are sometimes made to the individual child. However, care must be taken in reporting of scores when a "nonstandardized" administration of a test is given. As Barnett (1983) has stated: "Any alterations to the original materials or standardized procedures technically precludes use of established norms, except in an informal manner" (p. 178). There are problems with a lack of investigations of the effects of special accommodations on the

resulting scores or on their reliability and validity (APA, 1985).

Increasing the efficiency of the evaluation process, without forfeiting reliability and validity, is the goal in using computer-assisted testing. Recent developments in technology may assist in accomplishing these goals. For example, computer-assisted testing can involve an increase in: "(a) efficiency, (b) scoring efficiency, and (c) standardization of test interpretation and decrease in (a) scoring errors, (b) scoring time, (c) report writing time and (d) interpretative misjudgments" (McCullough & Wenck, 1984, p.429). Increased reliability was also found when test instruments were administered by computer vs. pencil and paper versions (Brown & Douglas, 1984). In addition, several positive features, after the initial outlay of funds for equipment, are a decrease in cost, savings of staff time and increased accuracy and information to guide instruction (Becker & Schur, 1986).

An additional potential benefit is favorable client reaction, as well as partial support for equivalency (Burke & Normand, 1987). Burke and Normand (1987) suggested that "approximately identical frequency distributions of test scores in which no change in examinee rank is observed between the conventional and computerized versions would provide more sound evidence of their equivalence" (p. 46).

Another issue to be considered before administering a computer version of a standardized test might be familiarization with equipment prior to testing. Providing for a system for error checking and correction is also important to ensure reliability. The

issue of the client's right to confidentiality by securing the discs is also mentioned (Burke & Normand, 1987). Additionally, that "cautionary statements in manuals and elsewhere regarding confidence in interpretations based on such test scores; forms of tests that are modified for people who have various handicapping conditions should generally be pilot tested on people who are similarly handicapped to check the appropriateness and feasibility of the modifications" (APA, 1985).

Computer-assisted assessment is still in a developmental stage. This is particularly true in assessment with severely impaired persons.

The purpose of this study is to examine the PPVT-R as it is normally administered, with a computer-assisted version which will rely on the examinee to touch a switch to indicate the correct response, rather than to indicate by pointing. Witt, Elliot, Gresham, and Kramer (1988) stated:

> If two stimuli are close together, a child who has a jerky hand movement may not be able to indicate reliably to which of the two stimuli he is pointing. Also, children become extremely adept at perceiving small cues from an examiner concerning correctness of a particular response and may switch quickly back and forth between alternatives. (p. 375)

It is proposed by the researcher that the computer-assisted version will be at least as sensitive as the standard version and may allow for improved responding in cases where there is motor impairment, attention problems, or greater fatigue due to physical involvement.

CHAPTER II

METHOD

Subjects

Twenty subjects were chosen for this study, based upon their educational eligibility as trainably mentally impaired, as determined by development at a rate approximately 3 to 4 1/2 half standard deviations below the mean. The age range was from 6 to 24, with the median being 17. Vision and hearing screenings revealed no significant losses which would interfere with participation in the study.

The students have the ability to follow directions and are behaviorally able to comply with the testing situation. None of the subjects was severely physically impaired, which could affect ability to reliably point to the pictures. Several of the students were speech and language impaired.

Informed consent forms signed by parent/guardian and student were obtained.

Setting

The testing was conducted at the school where the students regularly attend. Because all of the students have been previously evaluated within this setting, it is familiar, and generally reinforcing for them to be tested; that is, they enjoy the one to one situation. The same examiner administered all of the tests, with

two professional staff providing reliability checks. The students were tested in a small office, or in a larger conference room, both of which were familiar to them.

Apparatus/Materials

The test used in this study was the PPVT-R, form M (Dunn & Dunn, 1981) as this is most commonly used in this setting to evaluate receptive vocabulary. An Apple IIe computer was used to program the test. Reprints of the pictures were laminated so as to fit over the screen, allowing a space of 1 1/2 inches under each picture. The pictures were highlighted by an arrow beneath each one, which paused at each picture for three seconds, then moving to the next in a clockwise fashion, until the space bar was pressed, indicating that a selection had been made. Order of initial highlighting was randomized in the program in hope that perseveration would be decreased. The computer was equipped with a voice synthesizer which indicated the number of the item and "ready". It also was programmed to note the basal and ceilings and to automatically either loop back to create a basal or to end the test when a ceiling was reached. The raw score was recorded by the computer.

Five such possibilities were programmed to correspond with what a trained examiner would do when administering the test according to standard procedures, as described in Dunn and Dunn (1981).

Procedure

The students were randomly assigned to one of two treatments: The PPVT-R as it is normally administered, requiring a pointing response for the desired answer, or to the computer-assisted version in which a switch is closed in response. The subjects were given whichever treatment they had not already received two weeks following the first administration.

Interobserver agreement to check the integrity of the computer program was done. Two independent examiners checked the program for accuracy by doing several trial administrations, comparing the correct response with the computer's response using the first five possible "cases" in the administration manual. Point by point agreement of the individual test items was done with agreement divided by agreement + disagreement x 100. The results were r = 100%. During administration of the tests, reliability checks were done by comparing the two independent examiners' results: point by point agreement divided by agreement + disagreement x 100. Reliability checks were done with .23% of the total administrations with .99% agreement.

Treatment A, or the regular administration of the PPVT-R requires minimal directions regarding the student's response. The examiner simply explains that some pictures will be shown and asks for the student to point to, or indicate in some manner (gesture, yes, or no) which is the correct response. Treatment B, the computer-assisted version of the same test was demonstrated using

the sample items included in the test. An explanation was given: "See the arrow? If you push here (space bar with an "x" to highlight it), the arrow goes away". After mastery, the practice items were added and the students were allowed to practice the response until mastering it. For 10 of the students, this was accomplished within five minutes. For the remaining four students, 10 minutes was required for three, and 12 minutes for the fourth. Mastery consisted of correctly responding to the practice items approximately 80% of the time. For the four students who required a longer training time, switch training was stopped after approximately eight consecutive correct responses to avoid extending the testing time beyond their fatigue level. If the cause of the errors involved delayed responding, this was determined by asking the student to also point to the correct response. With both tests, at the outset, the student was urged to look at all pictures before pointing to the correct one.

Students were positively reinforced for attention to direction during training and care was taken to give no particular feedback for correct or incorrect answers. The tests were hand scored to insure accuracy, as the computer was not programmed to correct input errors, such as response delay errors, which occurred with many of the students.

CHAPTER III

RESULTS

The graph (see Appendix) represents the subjects by age, treatment, and order of administration. Six of the 20 subjects were unable to complete the task within the single sitting, which included switch training and familiarization with the apparatus. Those subjects were not excluded from the study in order to obtain information about their cognitive abilities, as well as their ability to follow directions and respond to the items.

Of the remaining 14 subjects, half received treatment A, or the manual version of the PPVT-R first, and the other half, treatment B, or the computer-assisted version to start. As shown on Table 1, the mean differences were as follows: For computer-assisted to manual, a difference of -1.73; for manual to computer-assisted, a difference of -3. For the second group, one subject (#3), who showed high distractability and fatigue before completing the test on the computer, showed scores with an 11 point difference. Without this score, this group's mean difference was -.33.

As the data indicate, age and ability to perform the task of closing a switch in response to a verbal stimulus and pictures seem related. Rather, something correlated with age for this group is most likely a factor. The one older student (#17) who would not perform the computer-assisted test is not only functioning in the low end of the TMI range, but could not or would not comply with the switch training, even to consistently touch the same key (the

Table 1

Mean Differences of PPVT-R Scores by Order of Administration

	Computer $\bar{X}$	Manual $\bar{X}$	Difference
Computer Administration First	61.29	63	-1.73
Manual Administration First	59	62	-3
Between Manual and Computer Administrations	62.5	60.14	-2.36

spacebar). Subject #9, who showed an 11 point difference, took 12 minutes to learn the task, and became fatigued before she reached the ceiling on the computer test. Subject #8, a 13 year old, began the switch training agreeably, then refused to try the task when the pictures were added. Subjects #1, 2, 4, and 5 were all highly distractable and became fatigued or noncompliant before mastering switch training, but complied with the manual administration. Subject #6, who showed a six point difference in favor of the computer-assisted test, was noted to give rapid responses during the standard administration, not always fully attending to all of the pictures before responding. The improved score on the computer-assisted version could have been due to the fact that the cursor directs the attention to all of the items.

Generally, the students reacted favorably to the computer task, enjoyed hearing the voice synthesizer announcing the item number, and were attentive and diligent in closing the switch. While some students performed easily within the three second interval, for

others it was too long, and for a few, not long enough. As earlier stated, having a cursor designate the item did serve to direct the attention for many students.

CHAPTER IV

DISCUSSION

As has been stated, the mean difference between the two groups was -1.73 points greater for the manual version of the test when it was administered first and three points greater for the manual administration when it was given second. For most of the students who were able to comply with the switch closing task within five minutes, the mean differences were negligible. With screening of students prior to this computer-assisted test to expedite the procedure and increase efficiency, the computer would give about the same results as the manual administration with this population.

Considerations when using a computer-assisted test are the time required to set up the program and train the students to close the switch. This, in the testing situation, added to the effort the student was required to put out, and used time the students needed to complete the task before becoming too fatigued to continue. Another consideration is that the stimulus duration, which in a normal testing situation could be geared to the individual student's needs, did not always "fit" the student, causing response lag errors. While these were not recorded as errors, they were sometimes distracting to the student. The three second interval also added to the time the test took to be administered, at times reducing efficiency by nearly half. It was noted that those who had a greater number of response lag errors also did not do as well with the computer-assisted task vs. the standard administration.

The issue of computer-naivete of the student is an important one for the examiner. Screening individuals for ability to perform switch closing tasks, or allowing for practice and training before actual testing takes place would save frustration for both the examiner and examinee, and ensure a "better fit" for the task at hand, and avoid task-fatigue effects. Assessment systems, now available from various manufacturers, give the examiner information about the client's capabilities, and allow for practice and training.

For those students who were unable to perform the task consistently, inability to sustain attention and difficulty with following two-step directions were observed. Another difficulty was poor visual tracking; these students would not consistently look at the screen when the stimulus word was presented, and seemed confounded by having to touch the spacebar at the same time, the task was too complex for them. A simple switch rather than the potentially confusing keyboard could be used. Overton and Scott (1972) found when requiring mentally retarded students to perform a similar task, more subjects were unable to obtain a basal score with the automated version than the manual one, and felt that this disadvantage could be eliminated if pretraining were extended. In comparing means for automated (visual display apparatus) versus manual version of the PPVT-R, they found nonsignificant differences.

The importance of efficiency in testing moderately to severely impaired clients cannot be overemphasized, and tailoring testing to reduce the length of time the test requires is an important

consideration. Garrison and Baumgarten (1986) found that tailored testing resulted in measurement efficiency when using computer adaptive testing with a group of entry level deaf college students. Students' attitudes toward their testing experiences also were measured, and were favorable. While tailored testing is in principle superior to the conventional written mode of assessment, empirical evaluation of adaptive testing has proven to be difficult (Garrison and Baumgarten, 1986).

It was also noted by the examiner that sitting next to the student at the keyboard precluded any eye pointing or facial expression cues that could be inadvertently given in a more directly facing presentation. The students would turn and look questioningly at the examiner at times, seeking feedback for responses. This situation could be disadvantageous in reverse, however, if the examiner was relying on this type of feedback from the student. While the use of psychometric software can potentially increase the reliability of assessment results, including fidelity of administration and alternate forms of reliability (Fifield, 1989), procedural alterations and nonstandardized administrations preclude reporting results in a formal manner.

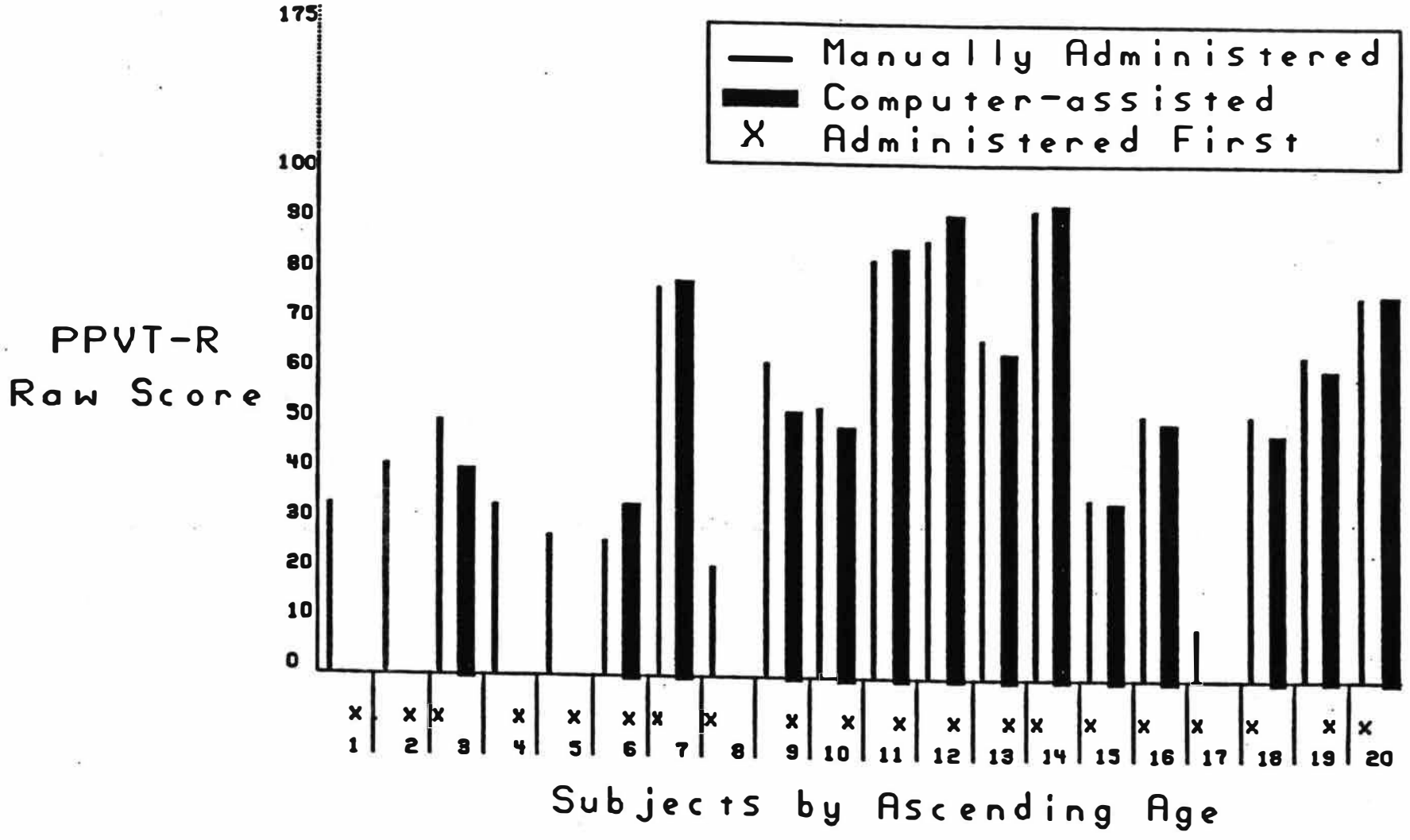
For further research, standardization of training procedures and identification of client characteristics which allows successful completion of switch closing tasks, as well as other more complicated tasks, is suggested. It should be reiterated, however, that due to the nature of various handicaps with special education students, to develop an accurate profile of strengths and

weaknesses, it is frequently necessary to assess skill areas using different approaches.

Another direction for development could be the use of this format with severely multiply impaired students, or physically impaired students. As with the TMI population, screening and training with a switch is suggested to determine whether it could be useful and efficient. As greater opportunities for learning are afforded the most handicapped of our population, as may be possible with computer-assisted programs, the potential for their growth and our ability to measure it, will be enhanced.

Appendix

PPVT-R Raw Scores


Manually Administered
Computer-assisted
X Administered First
PPVT-R
Raw Score
175
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
Subjects by Ascending Age

BIBLIOGRAPHY

American Psychological Association. (1985). Standards for educational and psychological testing. Washington, DC: Author.

Bagnato, S. J., & Niesworth, J. T. (1981). Linking developmental assessment and curricula. Rockville, MD: Aspen Systems.

Barnett, D. W. (1983). Procedures in the evaluation of children with low incidence handicapping conditions. In Nondiscriminatory multifactored assessment: A source book (pp. 162-182). New York: Human Sciences Press.

Becker, H., & Schur, S. (1986). The advantages of using microcomputer-based assessment with moderately and severely handicapped individuals. Journal of Special Education Technology, 8(2), 53-57.

Bracken, B. A., Prasse, D. P., & McCallum, R. S. (1984). Peabody Picture Vocabulary Test-Revised: An appraisal and review. School Psychology Review, 13(1), 49-59.

Browder, D. M., & Brookes, P. H. (1987). Assessment of individuals with severe handicaps: An applied behavioral approach to life skills assessment (pp. 177-178). Baltimore: Brooks.

Brown, & Douglas, T. (1984). Automated assessment systems in school and clinical psychology: Present status and future directions. School Psychology Review, 8(4), 455-460.

Brown, L., Falvey, M., Vincent, N., Johnson, F., Ferrara-Parish, P., & Gruenwald, L. (1980). Strategies for generating comprehensive, longitudinal, and chronological-age-appropriate individualized education programs for adolescent and young adult severely handicapped students. Journal of Special Education, 4, 199-216.

Burke, M. J., & Normand, J. (1987). Computerized psychological testing: Overview and critique. Professional Psychology: Research and Practice, 18(1), 42-51.

Carroll, J. B., & Horn, J. L. (1981). On the scientific basis of ability testing. American Psychologist, 36, 1012-1020.

Cruickshank, W. (1986). Disputable decisions in special education (p. 89). Ann Arbor: University of Michigan Press.

Davis, W. (1986). Resource guide to special education (2nd ed.). Boston: Allyn & Bacon.

Dunn, L.M. (1959). Peabody Picture Vocabulary Test (1st ed.). Circle Pines, MN: American Guidance Service.

Dunn, L. & Dunn, L. (1981). Peabody Picture Vocabulary Test-Revised (PPVT-R). Circle Pines, MN: American Guidance Service.

Fifield, M. B. (1989). Psychoeducational testing and the personal computer. Journal of Special Education Technology, 9(3), 136-143.

Garrison, W., & Baumgarten, B. (1986). An application of computer adaptive testing with communication handicapped examinees. Educational and Psychological Measurement, 46, 23-35.

Groeneweg, G., Conway, D. G., & Stan, E. A. (1986). Performance of adults with developmental handicaps on alternate forms of the Peabody Picture Vocabulary Test. Journal of Speech and Hearing Disorders, 51, 259-263.

Matey, C. (1985). Best practices in working with handicapped children. In Best practices in school psychology (p. 285). Kent, OH: Thomas & Grimes.

McCallum, R. S. (1985). Review of Peabody Picture Vocabulary Test. In The Ninth Mental Measurements Yearbook, 2 (pp. 1127-1129). Lincoln: University of Nebraska Press.

McCullough & Wenck. (1984). Current microcomputer Apple in school psychology. School Psychology Review, 13(4), 429-439.

Naglieri, J. A. (1982). Use of the WAIS-R and PPVT-R with mentally retarded children. Journal of Clinical Psychology, 38(3), 635-637.

Overton, G. W., & Scott, G. K. (1972). Automated and manual intelligence testing: Data on parallel forms of the Peabody Picture Vocabulary Test. American Journal of Mental Deficiency, 76(6), 639-643.

Prout, H. T., & Schwartz, J. F. (1984). Validity of the Peabody Picture Vocabulary Test-Revised with mentally retarded adults. Journal of Clinical Psychology, 40(2), 584-587.

Reckhase, M. D. (1986). Position paper on the potential use of computerized testing procedures for the national assessment of educational progress (pp. 2-20). American College Testing Service.

Sattler, J. M. (1988). Assessment of children (3rd ed.). San Diego: Author.

Switzsky, H., Rotatori, A. F., Miller, T., & Freagon, S. (1979). The developmental model and its implications for assessment and instruction for the severely/profoundly handicapped. Mental Retardation, 17(3), pp. 167-169.

Witt, J. C., Elliot, S. N., Gresham, F. M., & Kramer, J. J. (1988). Assessment of special children: Tests and the problem solving process (p. 375). Glenview, IL: Scott & Foresman.

Worthing, R. J., Phye, G. D., & Nunn, G. D. (1984). Equivalence and concurrent validity of PPVT-R forms L and M for school-age children with special needs. Psychology in the Schools, 21, 296-299.