The feasibility of administering some tests directly under computer management was investigated with Raven's Progressive Matrices Test (RPMT), a nonverbal test of general aptitude, chosen as the instrument for the study. Subjects were 76 deaf students in a vocationally-oriented postsecondary educational program. Half of the subjects were tested under conventional group procedures, and half under computer-managed conditions. Both groups were retested 12-28 days later under computer-managed conditions. Three major objectives were to determine: if performance was affected by mode of presentation; whether, under untimed conditions, time required to complete the test varied under the two conditions; and the coefficient of stability of the RPMT. Mode of presentation was found to be unrelated either to student performance of time taken to complete test, demonstrating that use of computers in testing does not affect test results and establishing the feasibility of using computers to administer, score, and report selected standardized tests. Evaluation of the RPMT showed it to be a useful backup test but not reliable enough a test on which to exclusively base educational decisions. (KW)
COMPUTER-MANAGED TESTING:
A FEASIBILITY STUDY WITH
DEAF STUDENTS

Fred J. Hitti, Roger L. Riffer
E. Ross Stuckless
July, 1971

Division of Research & Training
National Technical Institute for the Deaf
Rochester Institute of Technology
One Lomb Memorial Drive
Rochester, New York 14623

This report was developed in the course of
an agreement with the U. S. Department of
Health, Education, and Welfare.
1. Introduction

(a) Purpose of the Study

The National Technical Institute for the Deaf (NTID), established by Act of Congress in 1965, joined Rochester Institute of Technology in 1966. NTID's primary purpose is to provide postsecondary vocationally-oriented instruction for deaf students from across the nation, leading to successful employment.

Each entering NTID student is administered an extensive battery of tests designed to provide some measure of his aptitudes, interests, academic achievements, communication skills, and other dimensions important in counseling and constructing appropriate programs of study (Walter, 1969, 1971). In addition, his progress is carefully monitored, in part through retesting, as he proceeds through his program. It is vitally important that test results be analyzed and then interpreted to the students' counselors, academic advisors, and to the student, with a minimum of delay if they are to be useful in planning the student's academic program.

The large volume of standardized test administration activity by NTID personnel and the need for rapid "turn-around" time in test scoring and feedback prompted the investigators to explore the feasibility of administering some tests directly under computer management. This study is the result of that exploration.

Raven's Progressive Matrices was chosen as the instrument for this study for several reasons. Interest had already been expressed by several NTID personnel in administering this test as a backup for other psychological data available on students. As a nonverbal test of general aptitude, the Raven's lessens the confounding of deaf students' general intelligence with their language skills.

1 NTID's Computer Assisted Instruction Center includes an IBM 1500 Computer-Assisted Instruction system, featuring an IBM 1130 Central Processing unit and 12 IBM 1510 terminals with cathode ray tubes, image projector displays and both keyboard and lite pen response modes.
Moreover, the Raven's lends itself to investigation particularly well because of its relative ease of administration both individually and in groups. Also, it was relatively straightforward to program the computer to administer this test.²

(b) The Computer in Education

We have long been familiar with the impact of the computer upon business and industry. Until recently, however, the application of computer technology to the solution of problems in education was limited largely to the business office and to the routines of student scheduling and grade reporting.

Of late, we have seen applications of the computer extended to such educational concepts as vocational guidance and direct instruction (Educational Technology, 1970). Computer-assisted instruction has received particularly prominent attention in the last half decade. Lekam (1970) has reported that in 1967, fewer than 100 instructional programs were available for presentation in a computer mode, while in 1969 this number had increased to 910. Subject areas are increasingly varied, and the educational levels now encompass preschool through adult education.

Computer-assisted instruction is receiving attention also within education of the deaf. Several years ago, students at the Kendall School at Gallaudet College in Washington, D.C., began to receive some of their instruction in mathematics through terminals in the school linked by telephone line to a computer located at Stanford University, California (Behrens, 1969). In 1970, Stanford University received a federal grant enabling it to extend CAI to language instruction for deaf students. Also, NTID provides instruction in mathematics to

²Coursewriter II Program Language. The help of Dr. O. Dennis Barnes, Director of the NTID Computer Center, and Mr. Kenneth Snyder, who wrote and debugged the computer program, was indispensable and is much appreciated.
deaf students within its CAI facilities (Newton, 1969). Rathe (1968, 1969) has suggested that CAI holds considerable promise as an aid in instruction of deaf students but suggests that special educators must move with this technology rather than wait for general educators to do so.

(c) Application to Testing

Relatively little attention has been given to the possible role of the computer in educational test administration. Lekan (1970) reports that the Kuder Vocational Interest Inventory, the Minnesota Multiphasic Personality Inventory and the California Achievement Tests have been administered and scored by computer. Numerous non-standardized tests have also been programmed for computer-managed administration. However, a search of the literature reveals limited activity in exploratory application of the computer to testing.

Computer technology would seem to address itself well to several problems in educational testing:

1. Computer-Managed Testing (CMT) assures that a test is uniformly administered and scored, since it is independent of the individual test administrator and his testing "style".

2. CMT frees the educator-psychologist from the time-consuming, and sometimes onerous task of administering certain tests.

3. CMT permits rapid scoring and item analysis.

4. CMT can offer immediate feedback to the teacher and the student on the student's performance.

5. Test results can be readily stored with other student data in a master information system, and
quickly retrieved for student planning and other purposes.

6. CMT permits branching so that the source of a student's difficulties can be specifically identified, thereby offering unusual diagnostic potential.

Yet the applications of the computer to testing and to instruction remain, for the most part, hypothetical. Adaptable computer systems and the necessary support personnel are not immediately available to most educational institutions. More important, educators are justified in expecting more evidence than researchers can now provide that computers can contribute significantly to the testing process.

(d) Selection of the Raven's Progressive Matrices Test

The Raven's Progressive Matrices Test is untimed and can be administered either to individuals or to groups. Its tester's manual provides separate guidelines for administration under the two conditions, but the differences are minor. The test consists of 60 multiple-choice items. These items are presented in a booklet, one item per page. Each item consists of a design or matrix from which a piece has been removed. The student must select the correct missing part from among six to eight alternatives at the bottom of the page. There are five subtests, A through E, each consisting of 12 items. The subtests become increasingly more difficult.

The norms for this test are based on the scores of 5,857 British adults (3,665 militiamen and 2,192 civilians). Because males seem to perform somewhat better than females on this test, separate norms are provided for males and females. A table converts a raw score into a
percentile based on the age and sex of the student.

The Raven's Progressive Matrices Test has been used by psychologists in testing deaf students for many years. Since it is a non-verbal test, it largely avoids the interpretive problems inherent in verbal tests of intelligence when they are administered to students with verbal deficiencies (Vernon and Brown, 1964; Vernon, 1968). While most psychologists of deaf students consider scores on the performance scales of such individually administered tests as the Wechsler Intelligence Scale for Children (WISC) and the Wechsler Adult Intelligence Scale (WAIS) to yield more valid estimates of intelligence, the Raven's is often administered as a backup test. There is evidence that deaf students perform similarly to hearing students on the Raven's (Farrant, 1964).

Wechsler (Buros, 1949) has observed that the Raven's, while generally acceptable as a test of intelligence, has several limitations, among these being its low ceiling. Bortner (Buros, 1965), and Cruickshank and Johnson (1967) have criticized the test in terms of validity and reliability, but continue to recommend it as a non-verbal test of intelligence, particularly with handicapped groups.

2. THE PROBLEM

Most people would agree that testing is a very important dimension in the educational process. However, to be useful, tests must inform. Educational testing should contribute to the teacher's knowledge of the student, and to the student's knowledge of himself. In turn, this knowledge should contribute to a dynamic environment in which educational planning is synchronized with the changing needs of each student.

The computer was introduced into education to serve several functions. However, its possible usefulness in support of educational testing programs
has received little attention. The National Technical Institute for the Deaf is committed to the concept of flexibility in programming for its deaf students. This implies an obligation to develop a testing program that will contribute significantly to decision-making. Based on these needs and resources, this study was undertaken to explore the feasibility of computer-managed educational testing. More specifically, the objectives of the investigation were as follows:

(a) To determine if performance of deaf students on the Raven's Progressive Matrices Test is influenced by test administration under conditions of:
   (i) conventional group presentation,
   (ii) computer-managed presentation.

(b) To determine if, under untimed conditions, there is a difference in the time required by deaf students to complete the Raven's Progressive Matrices Test under conditions of:
   (i) conventional group presentation,
   (ii) computer-managed presentation.

(c) To determine the coefficient of stability of the Raven's Progressive Matrices Test for postsecondary deaf students.

3. PROCEDURE
   (a) Students Tested

The students participating in this investigation were all engaged in a special summer program offered by NTID. This program, called a Summer Vestibule Program, introduces incoming deaf students at NTID to college
life and affords an opportunity for them to sample various program areas. Extensive evaluation and counseling activities take place concurrently.

A total of 76 deaf students were enrolled in this program during July and August of 1970. All 76 participated in the investigation (48 males and 28 females). In general, the characteristics of these students are described in other NTID reports (Walter, 1970, in press). For the purpose of this investigation, these students were placed in two groups under a stratified random procedure whereby 24 males and 14 females were assigned to each of two experimental treatments. Sex was considered in assignment to groups since males tend to perform somewhat better on the Raven's than do females.

(b) Design of the investigation

The data for this investigation were gathered during a four week period. As previously indicated, 76 students were equally divided into two groups, numbering 38 students each. Group I was tested on the Raven's under a conventional group procedure. Group II was tested at the same time under computer-managed conditions. Both groups were retested 21 to 28 days later under computer-managed conditions. Of the original 76 students, 5 did not appear for retesting, limiting the number of retested students to 71.

The research design is represented schematically in figure 1.

<table>
<thead>
<tr>
<th>First testing</th>
<th>Second testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group I</td>
<td>CGP¹</td>
</tr>
<tr>
<td>Group II</td>
<td>CMT²</td>
</tr>
</tbody>
</table>

¹conventional group presentation
²computer-managed testing
Figure 1. Design of the investigation

(c) Conventional Group Presentation

Group I was administered the Raven's Progressive Matrices Test under conventional group procedures, as outlined in the test manual (Raven, 1960). To accommodate their study schedule, the 38 students in this group were divided into several subgroups, none of which numbered more than 8. The test thus was administered to the various subgroups at different times.

In deference to the communication handicap of the group, the conventional procedure was modified in the following ways:

(i) To assure that all subjects in Group I understood the test proceedings, a deaf student assistant, quite proficient both in simultaneous communication and in speech production, was assigned the task of giving directions to each subgroup. The student assistant, thoroughly briefed on the test directions, simultaneously spoke and signed the directions to each of the subgroups. In addition, a principal investigator was present at all testing to answer questions.

(ii) The manual directs the test administrator to make sure the students answer the first item correctly before they go further. The tester is then required to determine if the student has answered the first 5 items correctly. If the answers are incorrect, he is required to bring this to the student's attention and to review the test directions. This procedure, however, was modified so that, for the sake of efficiency, completion of the
first five items became part of the actual directions. This
departure from the manual does maintain Raven's intent and in
no way diminishes the validity of the test.

(iii) The test, according to the manual, is untimed. However,
for purposes of this investigation a stopwatch was (covertly) used
to record the elapsed time in minutes between the initiation of
the directions and the return of each student's answer sheet.

(d) Computer-managed Presentation

Group II was administered the Raven's Progressive Matrices
Test under computer-managed conditions. Both Groups I and II were
readministered the same test under computer-managed conditions three
to four weeks later.

Under computer-managed administration, each student was assigned
an individual console, activated by a staff member of the CAI center. (A
principal investigator was present to assist any subject who might have
difficulty.) Directions were then presented in graphic form on a cathode
ray tube. The subject was given a test booklet, and was instructed by the
computer when to turn its pages. He was also informed to record his answer
on a typewriter keyboard which was a component of the console. The elapsed
time from the student's "sign on" to his "sign off" was recorded by the
computer.

As soon as the student completed the test, his name, subtest scores,
total score, and elapsed time were printed by a teletypewriter in an adja-
cent room. However, students were not informed of their scores after the
first testing in order to prevent feedback from influencing their performance
Additional Similarities and Differences in the Two Presentations

Under both modes of presentation, the subjects used the regular test booklets and were instructed to answer every item. The essential differences in the presentations were (1) the means by which students were given directions and (2) the means by which students gave answers. Under the conventional group condition, each subject was given directions through speech and signs, whereas under the computer-managed condition, each subject was required to read directions which appeared graphically on a cathode ray tube. In each instance, a principal investigator was available to answer questions.

Under the conventional group condition, each subject was required to record his answer on a separate answer sheet. Under the computer-managed condition, each subject was required to press the numerical key corresponding to the answer he had selected. When he did this, a message on the cathode ray tube instructed him to consider the answer he had just given; if he wished to reconsider, he could press the numerical key corresponding to his new answer. If on the other hand, he remained satisfied with his first answer, he could press the space bar of the keyboard and proceed to the next item. Unlike the student being tested under the conventional group condition the student being tested under computer management was required to answer each question in serial order.

Figure 2 summarizes several of the similarities and differences in the two presentations.
## Presentation

<table>
<thead>
<tr>
<th>Similarities</th>
<th>Conventional</th>
<th>Computer-managed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raven's Progressive Matrices Test used, with same test booklets.</td>
<td>Principal investigator present to answer questions.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Differences</th>
<th>Tested in groups of 8</th>
<th>Tested individually, one student at a console</th>
</tr>
</thead>
<tbody>
<tr>
<td>Directions given by speech and simultaneous communication.</td>
<td></td>
<td>Directions presented graphically on cathode ray tube.</td>
</tr>
<tr>
<td>Time required to take test recorded by hand.</td>
<td></td>
<td>Time required to take test recorded by computer.</td>
</tr>
<tr>
<td>Student recorded answer on conventional answer sheet.</td>
<td></td>
<td>Student recorded answer on console keyboard.</td>
</tr>
<tr>
<td>Student permitted to answer questions in any order and to revise earlier answers.</td>
<td></td>
<td>Student required to answer questions in serial order with no opportunity for revision of earlier answers.</td>
</tr>
</tbody>
</table>

Figure 2. **Major similarities and differences in the two presentations.**
4. FINDINGS

(a) The First Objective

The first objective of the investigation was to determine whether the scores of deaf students on the Raven's Progressive Matrices Test differ under conventional group and computer-managed conditions. Table 1 indicates the means and standard deviations of the scores of the two groups under these two presentation modes at the initial testing.

Table 1

Means and Standard Deviations of Raw Scores

<table>
<thead>
<tr>
<th>Students</th>
<th>Group Presentation</th>
<th></th>
<th></th>
<th>Computer-Managed Presentation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M</td>
<td>SD</td>
<td>N</td>
<td>M</td>
</tr>
<tr>
<td>Male</td>
<td>24</td>
<td>52.04</td>
<td>3.68</td>
<td>24</td>
<td>51.67</td>
</tr>
<tr>
<td>Female</td>
<td>14</td>
<td>48.64</td>
<td>4.82</td>
<td>14</td>
<td>49.86</td>
</tr>
<tr>
<td>TOTAL</td>
<td>38</td>
<td>50.79</td>
<td>4.45</td>
<td>38</td>
<td>51.00</td>
</tr>
</tbody>
</table>

It is noted that the performance of males was slightly superior under the group presentation while the performance of females was slightly superior under the computer-managed presentation. When male and female scores were totaled for each of the two presentations, the mean favored the computer-managed presentation by 0.21 points. Also, the dispersion of scores, as indicated by the standard deviation, is larger for both males and females under the computer-managed condition. Further, under both presentation modes, the means of the male scores were higher than those of the females which is consistent with previous results using hearing males and females.

An analysis of variance was performed in order to determine (1) whether group and computer modes affected performance in a statistically significant manner; (2) whether males and females differed in their performance in a statistically significant manner; and (3) whether there
was an interaction between mode of presentation and sex. The results of this analysis are reported in Table 2.

Table 2

<table>
<thead>
<tr>
<th>Sources</th>
<th>F</th>
<th>Significance (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presentation</td>
<td>.1270</td>
<td>n.s.</td>
</tr>
<tr>
<td>Sex</td>
<td>4.8917</td>
<td>p.05*</td>
</tr>
<tr>
<td>Presentation X Sex</td>
<td>.4555</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

Table 2 reveals that no statistically significant difference occurred between the performance of those students who were administered the test under group procedures and the performance of those under computer-managed procedures. Deaf students apparently perform similarly under either condition.

Table 2 also indicates that male students performed significantly higher than female students. This is consistent with similar findings for other studies using the Raven's with hearing populations.

Finally, Table 2 points out that the performance of males and females was independent of the mode in which the test was presented. The higher scores for males cannot be attributed to a hypothetical superiority with mechanical devices which enabled them to adapt more quickly to the computer console.

(b) The Second Objective.

The second objective of the investigation was to determine whether the time taken by deaf students to proceed through the Raven's Progressive Matrices Test differed when they were presented the test under conventional
group and computer-managed conditions. Table 3 indicates the means and standard deviations of the time in minutes required by males and females under the two conditions.

Table 3

Means and Standard Deviations of Minutes Taken to Complete Test

<table>
<thead>
<tr>
<th>Students</th>
<th>Groups Presentation</th>
<th>Computer-Managed Presentation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M</td>
</tr>
<tr>
<td>Male</td>
<td>24</td>
<td>29.29</td>
</tr>
<tr>
<td>Female</td>
<td>14</td>
<td>26.57</td>
</tr>
<tr>
<td>TOTAL</td>
<td>38</td>
<td>28.29</td>
</tr>
</tbody>
</table>

Inspection of Table 3 reveals that the total group under conventional group conditions required 2.08 minutes less to take the test than the total group under computer-managed conditions. Females students took less time than male students to complete the test under both conditions. It is noteworthy also, that the dispersion of time required under both conditions, for both male and female students, was greater under computer-managed conditions than under group conditions (6.59 minutes versus 11.88 minutes).

Table 4 is a presentation of the results of an analysis of the variance in the time taken under the two presentations by the two sexes.

Table 4

Statistical Significance of Differences in Time Required to Complete Test by Presentation and by Sex

<table>
<thead>
<tr>
<th>Source</th>
<th>F</th>
<th>Significance (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presentation</td>
<td>.8655</td>
<td>n.s.</td>
</tr>
<tr>
<td>Sex</td>
<td>1.0504</td>
<td>n.s.</td>
</tr>
<tr>
<td>Presentation X Sex</td>
<td>.205</td>
<td>n.s.</td>
</tr>
</tbody>
</table>
Table 4 reveals no statistically significant differences in the time taken by males and females to complete the test, suggesting that the time taken by males and females to complete the test is independent of the manner in which it is presented.

A secondary interest of the investigators concerned the possible relationship between elapsed time to complete the test and test score. As a consequence, correlation coefficients between time and performance scores were calculated. No statistically significant relationships between the dimension of time and score were found.

(c) The Third Objective

The third objective of the investigation was to determine the coefficient of stability, or test-retest reliability, for the Raven's Progressive Matrices Test on a population of post-secondary deaf students under computer-managed conditions.

It should be recalled that half the subjects (38) were originally tested under group conditions and half under computer-managed conditions. Since no significant differences were found in the scores of these two groups, their scores can legitimately be pooled, thereby yielding a single group of 76 subjects. Of the 76, 71 were retested. The mean and standard deviation of these 71 retest scores are reported in Table 5, as are the mean and standard deviation of the original 76 scores.

Table 5

Mean Performance Scores and Standard Deviations of Original 76 students and 71 students retested 3-4 weeks later

<table>
<thead>
<tr>
<th>Performance Scores</th>
<th>Initial Test (Early in July)</th>
<th>Retest (3-4 weeks later)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>76</td>
<td>50.89</td>
<td>5.00</td>
</tr>
</tbody>
</table>
Table 5 reveals an average increase from test to retest of 1.83 points and a .97 decrease in the standard deviation.

As for the coefficient of test-retest reliability, the correlation between the first and second performance scores (for the 71 subjects who were tested twice) was .65, and the standard error of measurement was 2.37. A coefficient of correlation of .65 is relatively low for a test-retest situation with a standardized instrument. It compares quite unfavorably with the coefficient of .93 reported in the manual (Raven's, 1960) for the particular age group being tested. However, such a low coefficient may be due to the small variance in the subject's test scores. This lack of variance also explains the relatively low standard error of measurement.

The internal consistency of test performance was calculated using the Kuder-Richardson Formula 21. The coefficients of internal consistency are .70 and .62 for the performances of the 71 subjects over the two test settings. These findings suggest that neither the internal consistency nor the stability of the Raven's test, using a deaf student population, over a 3-4 week period is particularly high. At best, Raven's test may be of some use as a "backup" instrument for testing deaf students.

5. DISCUSSION

This investigation was undertaken to determine the feasibility of using the computer to administer, score, and report selected standardized tests. The results lend support to the concept.

Within the context of this study, the mode of presentation, that is, conventional group or computer-managed administration, was found to be unrelated to student performance. This would suggest that the introduction of the computer to the actual test administration neither interferes with nor enhances a student's test performance. This finding should serve to reduce concerns that the computer affects test performance.
Another dimension explored within this investigation was the relative time taken to complete an untimed test under conventional and computer managed conditions. Would the time lapse between beginning and ending a test be influenced by the mode of presentation? While the Raven's test is untimed, many standardized tests are timed, and scores can be influenced by the time available for a student to answer questions. In this investigation, the mode of presentation was found to be unrelated to the amount of time taken to complete the tests. A computer appears neither to speed up nor to slow down the rate at which students answer a test.

One attractive feature of a computer is its ability to record time accurately. A frequently discussed concept in psychological testing is that of latency. Does the time it takes a student to complete a test tell us something about his performance? Within this study, no relationship was found between test performance and the time taken to complete the test. Students who worked slowly were no more nor less likely to have high scores than students who worked quickly.

The third objective of the investigation was to examine the Raven's Test itself. How much confidence can one have in this test as a stable instrument? Does the score for a particular student mean anything? Findings based on the testing and retesting of 71 deaf post-secondary students suggest that individual students tend to score with relative consistency, but not with the consistency desirable in a test which is used to make educational decisions of major consequence. Essentially, the findings of this study are in agreement with recommendations of others: the Raven's Progressive Matrices Test can be a useful backup test, but not one on which significant educational decisions should be exclusively based (cf. Wechsler, Bortner, and Cruickshank and Johnson).
6. **POSTSCRIPT**

What did this investigation, in summation, really say? What relevance does it have for the education of students, more specifically for the education of deaf students? For the investigators, the results suggest that the computer can legitimately play a major role in the testing of deaf students as part of a total educational information system. The present costs of acquiring and maintaining a computer and the necessary costs of supporting computer personnel are substantial. A school could not justify these costs for testing alone. However, when testing is made part of an educational information system that includes the maintenance of educational information on students and the application of the computer's potential for instruction, then computer-managed testing warrants consideration. This investigation suggests an application for the computer in education as part of a comprehensive educational information system.


15. Vinsonhaler, J. E. et al.
An Experimental Study of Computer Aided Testing.
Information Systems Laboratory, Michigan State University, 1965.

16. Walter, G. G.,
Profile of Deaf Students Entering NTID in September, 1970.

17. Walter, G. G.,
Profile of Deaf Students Entering NTID in September, 1969.