R E P O R T R E S U M E S

ED 014 662 AC 001 625

EXPERIMENTAL EXPLORATIONS IN PROGRAMMED INSTRUCTION AND OBJECTIVE TESTING MEASURES, REPORT OF THE "VARIABLES INFLUENCING BEHAVIOR" PROJECT, PAPER 2.

BY- BERMAN, MARK L.
ARIZONA STATE UNIV., TEMPE

PUB DATE AUG 66

EDRS PRICE MF-$0.25 HC-$1.88 45P.

DESCRIPTORS- *EVALUATION, *INTELLIGENCE TESTS, *REINFORCEMENT, *TIMED TESTS, *READING TESTS, MEASUREMENT INSTRUMENTS, TESTING, RESEARCH, EXPERIMENTAL GROUPS, PARTICIPANT CHARACTERISTICS, MINORITY GROUPS, ENGLISH (SECOND LANGUAGE), OBJECTIVE TESTS, TEST RESULTS, PROGRAMED TESTS, ANALYSIS OF VARIANCE, EDUCATIONAL BACKGROUND, EDUCATIONAL DISADVANTAGEMENT, MEXICAN AMERICANS, AMERICAN INDIANS,

AS ONE PHASE OF RESEARCH IN APPLIED ANTHROPOLOGY, YAQUI INDIAN AND MEXICAN MEN IN ARIZONA HAVE PARTICIPATED FOR THREE YEARS IN EXPERIMENTAL PROGRAMED COURSES IN BASIC ENGLISH AND ARITHMETIC. THE STUDENTS HAD PREVIOUSLY HAD AN AVERAGE OF FIVE YEARS' FORMAL SCHOOLING. A BATTERY OF STANDARD OBJECTIVE TESTS WAS GIVEN TO MEASURE IQ AND ABILITY TO READ AND COMPREHEND ENGLISH. A FIRST SERIES OF TESTS WAS ADMINISTERED WITH SPECIFIED TIME LIMITS OBSERVED. A SECOND SERIES WAS GIVEN ALSO WITH THE TIME LIMITS. THEN STUDENTS WERE ALLOWED TO COMPLETE THE TESTS. THE CONCESSION OF TIME TO WORK TO COMPLETION CHANGED IQ PERCENTILE RANKS FROM 11 OR 12 TO 66, 77.5 AND 33. VARIOUS TESTS OF READING, VOCABULARY, AND OTHER SKILLS YIELDED HIGHLY INCONSISTENT EVALUATIONS OF THE STUDENTS' ABILITIES. A THIRD SERIES OF TESTS WAS BASED ON A NOVEL EXPERIMENTAL APPROACH TO PROGRAMED LEARNING. AN ADJUSTING SCHEDULE OF REINFORCEMENT FOR MEETING TIME AND ERROR CRITERIA WAS PUT IN FORCE. THIS INVOLVED RAISING OR LOWERING REQUIREMENTS FOR REINFORCEMENT USING THE STUDENT'S OWN PERFORMANCE AS A MEASURE. RESULTS INDICATED THAT BOTH SPEED AND ACCURACY CAN BE CONTROLLED BY REINFORCEMENT CONTINGENCIES. (ALSO INCLUDED ARE SIX REFERENCES AND 32 TABLES.) (AUTHOR)

Report of the "Variables Influencing Behavior" Project

Dr. J. A. Jones, Director

Arizona State University

Tempe, Arizona

Paper No. 2

EXPERIMENTAL EXPLORATIONS IN PROGRAMMED INSTRUCTION AND OBJECTIVE TESTING MEASURES

Mark Laurence Berman

Department of Psychology

Arizona State University

Tempe, Arizona

August, 1966

TABLE OF CONTENTS

ABSTRACT

As one phase of research in applied Anthropology, Yaqui Indian and Mexican men in Arizona have participated for three years in experimental programmed courses in basic arithmetic and English. The students had previously had an average of five years formal schooling. A battery of standard objective tests was given to measure IQ and also ability to read and comprehend English. A first series of tests was administered with specified time limits observed; a second series was given also with the time limits, then the students were allowed to complete the tests. The concession of time to work to completion changed IQ percentile ranks from 11 or 12 to 66, 77.5 and 63. Various tests of reading, vocabulary, etc., yielded highly inconsistent evaluations of the students' abilities. A third series of tests was based on a novel experimental approach to programmed learning; an adjusting schedule of reinforcement for meeting time and error criteria was put in force. This involved raising or lowering requirements for reinforcement using the student's own performance as a measure. Results indicated that both speed and accuracy can be controlled by reinforcement contingencies.

CHAPTER I

INTRODUCTION

Under Dr. J. A. Jones, Department of Anthropology, Arizona State University, a research project in "Variables Influencing Behavior in Indigenous Non-Western Peoples" began July 1, 1963. Activities have centered in Guadalupe, Arizona, a Mexican - Yaqui Indian community of about 3,000 people, situated eight miles south of Phoenix, Arizona.

The primary purpose of the V.I.B. Project is development of a new methodology for analyzing and influencing the behavior of persons living in small social units. We plan to develop procedures that will have cross-cultural application. Techniques proven successful in one society can then be adapted to other societies. In response to a strong interest expressed by the people in Guadalupe, educational activities have constituted one phase of the Project's research.

The data presented in this report stem from programmed courses in practical English, adapted to Yaqui and Mexican students, aged 17 to 40 years, who had failed to get adequate skill in the use of the English language during their school years. It became necessary that we have some means of measuring the standing of the students, both on an absolute and a relative basis, in terms of such variables as intelligence quotient, reading comprehension, spelling, vocabulary, word meaning, paragraph meaning, etc. With scores from standard tests that measure such variables, it might be possible to plot a student's progress (or lack of it) in gaining an ability to use English. Therefore, seven well-known and accepted tests were administered to the students, first under precisely the instructions that accompanied test booklets, and later with variations suggested by our work.

As reported in full detail in previous papers (Berman, 1964, 1965) a programmed mathematics course teaching fractions was developed and tested in 1963. This was followed by an English program, designated Level 3, constructed and tested by the author. It involved programmed presentation of elementary, 5th to 9th grade, English grammer and spelling. The student population included several men who had participated in the mathematics course.

Data from the Level 3 testing were analyzed, and certain aspects of the program were deleted or revised. In 1965, the author's Level 4 English program reviewed Level 3 material, then continued on, stressing reading speed and comprehension.

A Level 5 program, the only commercially produced material used in the English instruction described herein, covered grammar and syntax at an 8th to 11th grade level. This was tested for two months in early 1966.

This report presents performance figures from the Level 3 and Level 4 classwork, along with scores made by the students on the battery of objective tests given at the end of each course. The Level 5 program involved experimental use of an adjusting schedule of reinforcement for meeting time and error criteria.

Our results challenge the belief that an intelligence test is necessarily a test of a person's ability to solve problems. The figures also point to a serious bias that discriminates against students who simply cannot read the test questions fast enough, or who work through a body of material at a deliberate pace. This bias is of different weight and consequence in different tests, therefore it may be of crucial importance which test is given to a student who has an inadequate school or English language background. Our ex-

perimental work in the Level 5 course suggests that both speed and accuracy can be influenced by devices of programmed instruction and behavioral psychology. Classroom application of the adjusting schedule or equilibrium value technique would seem to offer immense possibilities for improving student performance.

CHAPTER II

PROCEDURE

Level 3 Procedure:

Seven students started this English program; their ages ranged from 21 to 40 years. Four students finished. A battery of standard tests was given these four at the end of the course. Test instructions for time limits were strictly enforced.

Level 4 Procedure:

Four students started and three finished this English program; their ages ranged from 30 to 40 years. All had participated in Level 3 instruction. The same battery of tests, using the second forms of each, was given at the end of the course. Time limit instructions were strictly observed, then after time had been called, the students were allowed to complete the test.

Level 5 Procedure:

Three students started and finished the Sullivan English program. Procedure and results will be discussed in a later chapter.

Tests Administered:

(1) Otis Group Intelligence Test, Advanced Examination, Forms B and A.

(2) Ohio Scholarship Tests, Techniques in Reading Comprehension for Junior High School, Grade 9, First and Second Every Pupil Tests.

(3) Ohio Scholarship Tests, Spelling and Vocabulary, Grade 6, First and Second Every Pupil Tests.

(4) Ohio Scholarship Tests, Elementary Reading - General Ability, Grade 6, First and Second Every Pupil Tests.

(5) Durrell-Sullivan Reading Achievement Test, Word Meaning and Paragraph Meaning, Intermediate Forms B and A.

(6) Stanford Achievement Test, Spelling and Language, Forms W and X.

(7) Metropolitan Achievement Tests, Intermediate Reading Test, Word Knowledge and Reading (Comprehension), Forms Bm and Am.

Note: Each of the above tests came with 2 forms The first form was administered after the Level 3 program, the second after the Level 4 program.

CHAPTER III

CLASSROOM DATA and DISCUSSION, LEVELS 3 and 4.

TABLE I. Level 3 performance Data:

Student	% Correct	Frames/Hour	Hours spent to finish program
1	92.5	50	30.25
2	92.6	39	37.75
3	95.2	45	33.25
4	97.1	54	27.50
*5	*90.8	*73	*20.75
*6	*96.4	*90	*16.67
*7	*89.5	*56	*26.78
Average:	94.0	51.57	**27.50

* Did not complete Level 3 Program
** Assuming Students 5,6,and 7 would have continued at same pace.

REMARKS

8,200 frames were completed, with a total of 493 errors; the total time spent was 158.76 hours. The range in time necessary to complete Level 3 (assuming that Students five, six, and seven would have continued at the same pace) was 16.75 to 37.75 hours.

Of the seven students, numbers six and seven were by far the youngest, being 23 and 21 years of age at the start of the program. These younger students worked much faster than the older men, averaging 76.1 frames per hour, as compared with 47.9 for the five older students. The average percent correct was approximately the same for both groups. A possible explanation of the younger students' greater speed is the fact that these two had dropped out of school only seven years before this class, while students one through five, all in the 30 to 40 year age bracket, had not been in school for an average of

twenty years. Therefore the routine of attending classes and meeting study requirements may have been less familiar to the older men and thus they worked more slowly.

TABLE 2. Level 4 performance Data:

Student	% Correct	Frames/Hour	Hours spent to finish program
1	97.7	51	8.63
2	92.0	43	10.23
3	95.9	55	8.00
4	98.6	54	8.15
Average:	94.0	50.75	8.75

REMARKS

1760 frames were completed, with a total of 106 errors; total time spent was 35 hours. The range in time necessary to complete Level 4 was 8.00 to 10.23 hours.

In three of four cases, the percentage correct on Level 4 increased over that made on Level 3. In the case of the one student who showed a decrease, this decrement was only .6 of one percent.

In every case but one, the speed of the students also increased, except in the case of student four, whose speed remained stable from Level 3 to Level 4.

The overall average percent correct on Level 4 was exactly the same as on Level 3, 94.0%. The average number of frames completed per hour was almost identical, being 51.57 for Level 3 and 50.75 for Level 4, a difference of only .82 of one percent.

(Without exception, there was an inverse relationship between the amount of speed increase and the amount of accuracy increase when comparing Level 4

with Level 5; to be discussed in the next section.

The student with the greatest speed increase showed the second-lowest accuracy increase. The student with the greatest accuracy increase showed the third-lowest speed increase. The student with the least speed increase showed the second-greatest accuracy increase. Finally, the student with the least accuracy increase showed the second-highest speed increase.

This agrees with "common sense," which suggests that speed increases can easily lead to accuracy decreases, while accuracy increases often come at the expense of speed.)

CHAPTER IV

OTIS GROUP INTELLIGENCE TEST DATA and DISCUSSION

TABLE 3. Advanced Examination, Form B (Administered after Level 3):

Student	IQ Score	Percentile Rank
1	88	16.00
2	85	11.00
3	82	6.45
*4	89	18.00
Average:	86	12.86

* Student 4 was killed before Level 4 program was completed.

TABLE 4. Advanced Examination, Form A (Administered after Level 4):

Student	IQ Score	Per. Rank	IQ Score with time for completion	Per.Rank with time for comp.
1	85	11.00	105	66.00
2	86	12.00	109	77.50
3	85	11.00	104	63.00
Average:	85.33	11.33	106	68.83

REMARKS

When the students were allowed time to complete the test form (A) their scores, averaged, put them in the top 31.17% of persons 19 years of age and over, as listed in the Otis scoring manual, rather than in the bottom 13%. One student placed in the top 22.5%. These scores were not surprising to the research staff whose subjective evaluation of the students had easily classified them as "bright," "alert," "intelligent."

Upon retesting, the results from Form A as compared with Form B showed little variation in the average score. However, individual scores showed

slightly more fluctuation. When the students were allowed to go to completion on Form A, a remarkable increase occurred, averaging 20.67 points on the IQ scale and 57.50 points on the percentile rank scale.

TABLE 5. Scores from Form A, with time limit enforced.

Student	Correct	Incorrect	Omitted	Percent correct of those answered
1	86	37	107	69.9
2	90	23	117	79.6
3	86	47	97	64.6
Average:				71.3

TABLE 6. Scores from Form A, with no time limit.

Student	Correct	Incorrect	% Correct of those previously omitted	% Correct of all questions
1	143	87	53.2	61.7
2	155	75	55.5	67.3
3	142	88	57.7	61.7
Average:			55.4	63.6

TABLE 7. Rank order of students.

Student	Rank on "A" with time limit	Rank on "A" without time limit	Rank on "A" answered correctly
1	2 (tie)	2	2
2	1	1	1
3	2 (tie)	3	3

TABLE 8.

Student	% of Questions answered correctly	% Correct to Completion
1	70	53
2	80	56
3	65	58

REMARKS

These figures damage the myth that questions missed on the intelligence test would still be missed if additional time were allowed, i.e., if students could work to completion. Students were correct on 55.4% of answers to questions that they could not complete within the time limit, compared with 71.3% correct on questions answered within the time limit.

TABLE 9. Scores from Form B, with time limit.

Student	Correct	Incorrect	Omitted	Percent correct of those answered
1	95	32	103	66.67
2	85	40	105	68.00
3	74	37	119	74.80
4	98	21	111	82.40
Average:				72.97

CHAPTER V

RELATION BETWEEN OTIS IQ RANKING AND SPEED AND ACCURACY RANKING IN PERFORMANCE ON LEVELS 3 AND 4

TABLE 10. Otis ranking and speed rankings.

A. Student	Form B Otis Ranking	Speed on Level 3 with time limit
1	2	2
2	3	4
3	4	3
4	1	1

B. Student	Form A Otis Ranking	Speed on Level 4 with time limit
1	2 (tie)	2
2	1	3
3	2 (tie)	1

C. Student	Form A Otis Ranking "to completion"	Speed on Level 4 with no time limit
1	2	2
2	1	3
3	3	1

REMARKS

There is a perfect correlation between IQ ranking on the Otis Form B and speed ranking on Level 3 in two of the four students' scores. For the other two students, the correlation is only one rank away. That is, the student ranked third in IQ was ranked fourth in speed, and vice versa.

There is a perfect correlation between IQ ranking on Otis Form A and speed ranking on Level 4 in one of the three students' scores. One score was

one rank away, while the third was two ranks away.

When students completed Otis Form A, there is a perfect correlation between IQ ranking and speed on Level 4 in one case. The other two students' scores show the maximum possible deviation, two ranks.

TABLE 11. Spearman Rank-Order Correlation Coefficients.

IQ/Speed Data	Coefficient
Table 9 A	Plus .80
Table 9 B	Minus .25
Table 9 C	Minus 1.0
Combining A & B	Plus .875
Combining A, B,& C	Plus .909

TABLE 12. Rank Difference Between IQ Rank and Speed Ranks, Level 3 and 4 From Table 9 A, B, and C.

Zero	One	Two	Three
4	3	3	0

REMARKS:

It appears that a strong positive relation exists between speed on both levels of the English program and scores on the Otis Group Intelligence Test. Observing Table 10 A, B, and C, we note that in every case where speed ranking on Level 4 declined relative to speed ranking on Level 3, the IQ ranking also declined on Form A relative to Form B. Also, in every case we note that where speed ranking increased from Level 3 to Level 4, the IQ ranking also increased. Finally, in the case of the one student whose speed ranking remained the same, the IQ ranking also remained the same.

These relations do not pertain with regard to a comparison of accuracy rankings on Levels 3 and 4 with IQ rankings.

TABLE 13. Otis ranking and accuracy rankings.

A. Student	Form B Otis Ranking	Accuracy on Level 3 with time limit
1	2	1
2	3	2
3	4	3
4	1	4

B Student	Form A Otis Ranking	Accuracy on Level 4 with time limit
1	2 (tie)	1
2	1	3
3	2 (tie)	2

C. Student	Form A Otis Ranking "to completion"	Accuracy on Level 4 with no time limit
1	2	1
2	1	3
3	3	2

REMARKS:

Comparing Otis Form B ranking to accuracy ranking on Level 3, we find that in no case is there any correlation between the two rankings. This contrasts with the situation found when comparing IQ and speed for the same form and level. In the latter, there were two perfect correlations out of four cases.

In the case of accuracy, three of the four cases show differences of one rank between the two measures. The fourth student has a three-rank difference of more than one rank between the two measures.

When comparing Otis ranking on Form A to accuracy ranking on Level 4 the same general lack of positive correlation prevails though the differences are somewhat less. We find one case of perfect correlation, one of one rank difference, and one of two rank difference.

Again, comparing Otis ranking on Form A to completion with accuracy ranking on level 4, we find two cases of one rank difference and a third case of two rank difference

TABLE 14. Spearman Rank-Order Correlation Coefficients.

10/ Accuracy Data	Coefficient
Table 12 A	Minus .20
Table 12 B	Minus .25
Table 12 C	Minus .50
Combining A & B	Plus .696
Combining A, B, & C	Plus .861

TABLE 15. Rank Difference Between IQ Rank and Accuracy Rankings, Levels 3 and 4, From Table 13 A, B, and C.

Zero	One	Two	Three
1	6	2	1

Comparing the data in the two tables of Spearman Correlation Coefficients, Tables 10 and 13, it is quite apparent that there is much less correlation between IQ rank and accuracy on levels 3 and 4 than there is between IQ rank and speed on levels 3 and 4.

The data (particularly, A and B of Table 10) strongly suggest that a functional relationship exists between the speed at which a person works through some material, eg., the English program, and the score he will attain on a test which requires him to complete the test within a specified time.

Other information (particularly, Tables 4, 5, and 6) lends credence to the strong possibility that it is not how accurately a person works, but how fast a person works that strongly effects his score on the Otis Group Intelligence Test.

It is of interest that the Otis tests apparently have a definite bias against the so-called culturally-disadvantaged person. Such persons, lacking effective reading skills (our students had, on the average, only five years of formal education), are heavily selected against due to the great emphasis on speed in the Otis Group Intelligence Test.

"Intelligence" is sometimes defined as the ability to solve problems, with "high intelligence" defined as the ability to solve difficult problems. For example, let us postulate three problems, one of which is very simple, another which is of moderate difficulty, and a third which is extremely difficult to solve. A person who can solve only the first problem might be called "mentally deficient." A person who could solve the first and second problems, but not the third, might be called a person of "average intelligence." The person who could solve all three problems might then be classified as a person of "high intelligence."

Suppose these three persons were put into a room and presented the three problems. Further suppose that the instructions were to work on the problems in the order in which they appeared on a typed piece of paper; that is, first, the very simple problem, second, the problem of moderate difficulty, and, third, the extremely difficult problem. As one of the conditions there is a limit placed on the amount of time the students have to work on the three problems. Let us say this limit is fifteen minutes.

The timer starts and the three persons being tested commence their work

on the problems. At the end of five minutes one of the people has finished the first problem. At the end of ten minutes this person is midway through the second problem, while another person is just starting to work on the second problem, and the third person is still working on the first problem.

At the end of the fifteenth minute, time is called. The three persons have progressed as follows: One person has finished only the first problem and is just beginning on the second problem; a second person is halfway through the solution of the second problem; the third person is just about to finish the solution to the second problem.

The answers to the problems are now scored. Each of the three persons completed the first problem and got the correct solution. Since no one completed the second problem, all three persons got credit for one correct answer out of a possible three.

Now the person administering the test notifies the three persons that he will allow each of them ten minutes more to complete the test. Following the additional ten minutes we find the following: the first person is still working on problem number two; the second person has answered problem number two, but incorrectly; the third person has completed problems two and three correctly.

Now we have some sort of grouping of the students: The first person has answered one problem correctly and none incorrectly; the second person has answered problem one correctly and problem two incorrectly; the third person has answered all three problems correctly. Therefore, we have a tie for second place, since two students have gotten a single problem correct.

Assume the examiner is very benevolent and allows the first and second persons another ten minutes to complete the test.

Following the additional ten-minute period the results are as follows: the first person has completed all three problems, getting two correct; the second person has completed all three problems, getting one correct; the third person has completed all three problems, getting a perfect score.

The problem arises of which is the "true" measure of the "intelligence" of these three people? Is it their score after fifteen minutes, after twenty-five minutes, or after thirty-five minutes?

If we take fifteen minutes as the baseline we find a three-way tie, each person getting one correct solution. Taking twenty-five minutes as the baseline we find a two-way tie for second place. Using thirty-five minutes as a baseline we find a clear-cut differentiation between the three persons with regard to their success in solving the three problems.

The situation fictitiously presented here is quite analogous to that which exists in the so-called "real world." By imposing artificial time limits on such tests as the Otis Group Intelligence Test, it is difficult to arrive at a "true" picture of the persons being tested, since each person works at his own speed.

If what we desire is a measure of how fast a person can answer questions, then we should change the names of many of our tests, for example, the Otis would become the Otis Group Quick Answers Test, or something similar.

Particularly penalized on such tests are those persons with slow reading habits, deliberate (i.e., slow-moving) problem-solving sytles, culturally-disadvantaged, etc.

CHAPTER VI

SCORES ON AND INTERPRETATIONS OF OBJECTIVE TESTS

TABLE 16. Percentile Scores, Ohio Scholarship Tests, Techniques in Reading Comprehension for Junior High School, Grade 9, First Every Pupil Test:

Student	Percentile
1	65
2	30
3	3
4	40
Average:	34.5

TABLE 17. Second Every Pupil Test:

Student	Percentile (time)	Percentile (to Comp.)
1	4	58
2	1	42
3	1	18
Average:	2.0	39.3

REMARKS

Overall, scores on the First Every Pupil Test in Reading (comprehension) were relatively high, considering that a ninth grade criterion was in effect. Only one student did very poorly here. On the Second Every Pupil Test in Reading when students were allowed to work to completion did they show a fair degree of competence. Even here one student still scored in the 18th percentile. The reasons for the great decrease in scores on the retesting are not clear.

Percentile scores decreased markedly in the retesting, going from 34.5 to 2.0, a decrease of 32.5 points. When allowed to go to completion, the to-

completion scores averaged only 4.8 points higher than the average scores on the initial testing. Comparing scores with the time limit in effect to scores going to completion, we find large increments. These were from 1.0 to 42.0, 4.0 to 58.0, and 1.0 to 18.0 respectively, or an average of over nineteen times the lower score!

TABLE 18. Percentile Scores, Ohio Scholarship Tests, Spelling and Vocabulary, Grade 6, First Every Pupil Test:

Student	Percentile
1	35
2	35
3	5
4	15
Average:	22.5

TABLE 19. Second Every Pupil Test:

Student	Percentile (time)	Percentile (to Comp.)
1	40	65
2	4	20
3	15	33
Average:	29.6	39.3

REMARKS

On this test the scores show very poor performance. Upon retesting there was a small increase in the average percentile score, 7.1 points. There was a further increase of 9.7 points when students were allowed to go to completion. On retesting, with students working within the time limit, the average score was below the 30th percentile for Grade 6.

TABLE 20. Percentile Scores, Ohio Scholarship Tests, Elementary Reading-General Ability, Grade 6, First Every Pupil Test:

Student	Percentile
1	30
2	25
3	1
4	35
Average:	22.7

TABLE 21. Second Every Pupil Test:

Student	Percentile (time limit)	Percentile (to Comp.)
1	50	85
3	4	12
Average	27.0	48.5

REMARKS

Apparently this particular test is one in which the students benefitted greatly from being allowed to work to completion.

Scores on this test, while higher than those received on the Spelling and Vocabulary test, are still relatively low.

There was a small increase in scores upon retesting from 2.7 to 27.0. When allowed to go to completion on the second test, there was a large increase, amounting to 21.5 percentile points.

TABLE 22. Grade Equivalents, Durrell-Sullivan Reading Achievement Test, Intermediate Form B:

Student	Word Meaning	Paragraph Meaning	Total
1	4.6	6.8	5.5
2	5.7	5.5	5.6
3	6.2	6.3	6.3
4	7.0	7.4	7.3
Average:	5.87	6.52	6.17

TABLE 23. Intermediate Form A.

Student	Word Meaning (Time)	Meaning (Time)	Total	Word Meaning (Completion)	Meaning (Comp.)	Total
1	6.8	6.6	6.8	7.1	9.2	8.4
2	5.2	4.9	5.1	7.1	9.4	8.4
3	6.4	4.9	5.9	7.1	8.2	8.1
Average:	6.13	5.46	5.93	7.10	8.93	8.30

REMARKS..

Upon retesting grade equivalent scores for Word Meaning increased .26 grades on the average. Scores for Paragraph Meaning increased 1.06. Total scores, however, decreased .24 grades on retesting. On Form A, when allowed to work to completion, all measures (Word Meaning, Paragraph Meaning, and Total Score) showed substantial increases over time-limit scores, being .9/, 3.4/, and 2.3/ grades respectively.

Scores on all three measures were higher for Form A to completion than corresponding scores for Form B with a time limit.

If the "to completion" scores are omitted, there were no large differences between Word Meaning scores (average of 6.00 grade level for Forms A and

B combined).

TABLE 24. Percentile Scores, Stanford Achievement Test, Spelling and Language, Intermediate I, Middle Grade 5 Baseline, Form W:

Student	Spelling	Language
1	96	38
2	74	22
3	92	8
4	53	28
Average	78.7	24.0

TABLE 25. Form X:

Student	Spelling (time)	Spelling (to Comp.)	Language (time)	Language (to Comp.)
1	99	99	46	48
2	7	60	24	54
3	88	88	12	32
Average:	64.6	79.0	27.3	44.6

REMARKS

When students are allowed to work to completion, increases in percentile scores may or may not occur, depending on the type of skill being measured. Apparently the Stanford Test puts much less emphasis on speed than does the Metropolitan Achievement Test (data in following section). There are some percentile points gained from going to completion on the Stanford Test. However, the magnitude of gain is small when compared to the gains achieved on the Metropolitan Test, Reading section, which shows a gain of 49.0 percentile points.

In administering the Stanford Test a second time there was a decrease

in the average Spelling percentile score of 14.1 points. In the case of Language, there was an increase of 3.3 points. When allowed to go to completion on Form X, the Spelling score increased 14.4 points, while the Language score increased 17,3 points.

The gap of approximately thirty-five to thirty-seven percentile points between the Spelling and Language scores was maintained, even after the students were allowed to work to completion on Form X.

This contrasts with the situation found for the Metropolitan Reading Achievement Test, in which the gap between Word Knowledge and Reading scores was completely erased after students were allowed to work to completion. Prior to allowing students to work to completion, the average difference was 41.3 percentile points.

TABLE 26. Grade Equivalents, Metropolitan Achievement Tests, Intermediate Reading Test, Form Bm:

Student	Word Knowledge	Reading (Comprehension)	Total
1	10 plus	8.1	
2	8.7	5.1	
3	8.7	5.7	
4	10 plus	7.3	
Average	9.35	6.85	8.10

TABLE 27. Form Am:

Student	Word (time)	Word (to Comp.)	Total	Reading (time)	Reading (to Comp.)	Total
1	10 plus	10 plus		6.8	10 plus	
2	7.9	9.2		4.2	7.3	
3	6.6	7.1		5.1	9.7	
Average:	8.17	8.77	6.77	5.36	9.0	8.89

REMARKS

The scores on Metropolitan Achievement Test are relatively high, far exceeding the same student's scores on any other tests. This seems to underline the fact that which test a student takes is all-important, and to stress the point of not relying on a single test as an accurate measuring device.

On Form Bm there is a large discrepancy between grade equivalents for the Word Knowledge section and the Reading Section; the difference is 3.5 grade levels, in favor of Word Knowledge.

On Form Am (time limit), the corresponding scores are 8.17 and 5.35, a difference of 2.81 grades, again in favor of Word Knowledge.

However, when students were allowed to work to completion on Form Am, the difference between Word Knowledge and Reading was completely erased, with the latter even showing .23 grade levels higher than the former. Large increments in Reading scores were obtained by allowing students to work to completion in that particular section, gains being, on the average, 4.64 grade levels.

On Form Am with time limit, (retesting) scores decreased 1.18 grade levels for Word Knowledge and 1.49 grade levels for Reading.

CHAPTER VII

CLASSROOM DATA AND DISCUSSION, LEVEL 5.

The level 5 program involved a novel experimental approach to programmed instruction, involving the use of "adjusting schedules of reinforcement," applied to individual student performance.

The procedure was as follows: Programmed English, written by M. W. Sullivan and published by the MacMillan Company was used for this class. This program, though not originally intended for written responding, was easily adapted for our procedure simply by changing the instructions to the students.

The program was arbitrarily divided into one hundred sections of approximately eighteen frames each. On the average, about fifty-four responses were called for in each section, since each frame required about three written responses.

A further division was made. The program was laid out in four separate parts of twenty-five sections each. The first part involved a particular set of reinforcement contingencies, the second part another set, and so on.

The adjusting schedules involved raising and/or lowering the requirements for reinforcement, using the student's own performance as a measure. If the student met the criteria for reinforcement, the requirements were systematically raised. If the student did not meet the criteria, the requirements were lowered. The requirements for reinforcement therefore adjusted themselves, depending solely on the student's performance.

Reinforcement involved receiving five points for meeting one of the criteria and ten points for meeting both criteria, when both contingencies were in effect.

Additional reinforcement involved the approval of the instructor following successful, i.e., criterion-meeting performance. As in previous levels of the overall English program, this approval was avidly sought by the students.

Two kinds of criteria were used, one based on time, the other based on the number of errors made per section. If the student finished the particular section he was working on in the same or less time than the criterion given, he received five points. If he made the same or less errors than the criterion, he also received five points.

A record of each student's performance was kept in a log. Point totals were kept, both on a section by section and a cumulative basis. Students could ask to see their point totals, but never did so. It was discovered that they were maintaining an approximate accounting of their own (and other's) point totals in their heads.

The criterion point changed each time a section was completed, the time or error allowances being raised or lowered depending on the student's performance.

If the student met the time criterion, he was given five minutes less for the next section. If he failed to meet the time criterion, he was given five minutes more for the next section.

If the student made the same or fewer errors than the maximum allowed for reinforcement, then on the next section the error allowance was decreased one error. If the student failed to meet the error allowance, the error allowance was raised one error.

At the beginning of the Level 5 program, time and error allowances were set arbitrarily for the students. Each student had the same beginning

allowances, time-80 minutes, and errors-10. Therefore, if the student completed the first section of frames in eighty minutes or less he received five points, and the allowance on the next section was lowered to 75 minutes.

If the student did not meet the time allowance and finished, for example, in 100 minutes, he was allowed five minutes more to complete the next section, which would be Section Two. The student would receive reinforcement, i.e., five points, if he completed Section Two in 85 minutes or less.

Similarly, if the student made ten errors or less in section one, he then received five points and the error allowance on section two was lowered to nine errors, a drop of one error. If the student made, for example, fifteen errors on the first section, he received no points. Then, in the following section, his error allowance was raised from ten to eleven. Then, if the student made eleven errors or less in the second section he would receive five points.

During the four parts of the program, the following contingencies were in effect:

Part 1: both time and error
Part 2: time only
Part 3: error only
Part 4: student's choice (he could select either time or error as the contingency)

TABLE 8. Data Level 5, Time.

Student	Section	Average Time (minutes)	Ave. Equilib. Value	Difference Avg. Time/EV	% Difference Avg. Time/EV
1	First 11	44.80 min	59.50 min.	- 14.70	- 24.79%
	Second 11	32.60	38.10	- 5.50	- 14.44%
	First 25	38.68	46.00	- 7.32	- 15.91%
	Second 25	24.08	25.00	- .92	- 6.68%
2	First 11	51.20	61.30	- 10.10	- 16.48%
	Next 7	47.60	55.70	- 8.10	- 14.55%

TABLE 28. (Cont'd)

Student	Sections	Average Time (minutes)	Ave. Equilib. Value	Difference Avg. Time/EV	% Difference Avg. Time/EV
	First 25	46.16	52.40	- 6.24	- 11.91%
	Next 18	29.55	26.67	+ 2.88	+ 11.08%
3	First 11	35.00	55.00	- 20.00	- 63.63%
	Second 11	40.80	36.30	+ 4.50	+ 11.24%
	First 25	37.20	52.40	- 15.20	- 29.01%
	Next 9	28.55	31.67	- 3.12	- 9.85%

TABLE 29. Data Level 5, Errors.

Student	Sections	Average Errors	Ave. Equilib. Value	Difference Avg.Errors/EV	% Difference Avg.Errors/EV
1	First 11	3.10	5.90	- 2.80	- 47.46%
	Second 11	1.50	1.30	+ .20	+ 15.40%
	First 25	2.40	3.36	- .96	- 28.57%
	Second 25	4.96	None		
2	First 11	6.10	6.50	- .40	- 6.15%
	Second 11	3.18	2.73	+ .45	+ 16.40%
	First 25	5.48	5.68	- .20	- 3.44%
	Next 18	4.40	None		
3	First 11	4.10	6.30	- 2.20	- 34.77%
	Second 11	3.20	2.70	+ .50	+ 18.50%
	First 25	3.56	4.20	- .64	- 15.24%
	Next 9	3.79	None		

REMARKS

The above figures show, though not conclusively, that students had slightly more trouble meeting the error allowance than the time allowance.

For the second 11 sections, all records show an inability to bring the number of errors below the error EV, or maximum error allowance. In the first 11 sections, however, all students were able to make fewer errors than the EV value, and thus were "reinforced."

Another picture arises with respect to time performance. For the second 11 sections, two out of three students completed these sections in less than the EV value. As with the error contingency, in the first 11 sections all students were able to meet the EV contingency.

While students had progressively more difficulty in meeting both the time and error contingencies set up, it was much more difficult for them to meet the error criterion as they moved through the first part (of twenty-five sections) of the Level 5 program.

Due to the nature of the reinforcement and contingency schedule arranged for the second part of the Level 5 program, no error EVs were maintained during this part. A time-only contingency was in effect. It is, however, possible to compute what the error EVs might have been if the error contingency had not been dropped during sections twenty-six to fifty. (For example, Student One's error EV at the end of section twenty-five was three errors. On section number twenty-six, the first section under time-only contingencies, Student One made two errors. If the error EV had still been in effect there would have been a reduction of the error EV to two errors, since meeting the error EV automatically caused it to be reduced by one error.)

Thus the possible error EVs can be plotted for sections twenty-six to fifty, and we calculate the following:

TABLE 30.

Student	Average Errors per Section		EV	Difference between Avg. and "EV"		
1	2nd 25 sections	4.96	4.00	(plus)	0.96	(24.0%)
2	sec. 26 to 43	4.40	4.00	(plus)	0.40	(10.0%)
3	sec. 26 to 34	3.79	3.33	(plus)	0.46	(13.8%)

REMARKS

Students continue to make more errors than the hypothetical EV for errors. Allowing for the fact that students completed different numbers of sections under time-only contingencies, we find that the average difference between the average number of errors made during time-only contingencies and the hypothetical error EV is 19.5%. This is compared to 15.8% for performance during the first twenty-five sections under time-and-error contingencies.

Allowing for the manner in which the error EVs were established for performance under time-only contingencies, we see that as might be predicted, the same difficulty in meeting error EVs prevails as did in sections 12 to 22 of the program.

TABLE 31.

Student	Ave. Time 1st 25 Sections	Time Next --	Time Next - -
1	38.63 minutes	24.08 (25 sec.)	[illegible].00 (1 sec.)
2	46.16 minutes	29.55 (18 sec.)	-
3	37.20 minutes	28.55 (9 sec.)	-

TABLE 32.

Student	Ave. Errors 1st 25 Sections	Errors Next --	Errors Next --
1	2.40	4.96 (25 sec.)	4.00 (1 sec.)
2	5.48	4.40 (18 sec.)	-
3	3.56	3.79 (9 sec.)	-

TABLE 33.

(a) Average time with both time and error contingencies - 40.75;

(b) Average time with time-only contingency - 27.10;

(c) Average errors with both time and error contingencies - 3.81;

(d) Average errors with time-only contingency - 4.56.

REMARKS

When the time-only contingency was in effect, the time taken dropped 33.5% and the number of errors rose 19.4%. This agrees with the predicted effect of having a time-only contingency following both time and error contingencies. The effect on time taken was greater than on the number of errors made. This can be explained by the probability that it is very easy to adjust the time one takes to complete a section of the program. A person can breeze through as fast as he can write.

On the other hand, he can go as slowly as he wishes. In regard to error rate, however, the lower limit to the number of errors made is not so strictly up to the whim of the student, unless the material is so simple that the student can easily get all answers correct.

If the material is difficult, it is relatively easy to make many errors but relatively hard to make few or no errors. In this sense the error rate is not as manipulable as the time rate and, hence, will probably not vary as much

as the time rate, when error contingencies are changed.

It appears that the initiation of time-only contingencies from sections twenty-six to fifty caused a precipitous rise in errors and a somewhat less-drastic drop in time taken. This is seen from the fact that, for Student One, errors on the first twelve sections averaged 2.92 per section, while errors on the next thirteen sections averaged 2.08 per section, a drop of 0.84 errors per section.

However, beginning on section twenty-six, when the time-only contingency was introduced, errors rose from the 2.40 average on the preceding twenty-five sections, to 4.96 on the next twenty-five sections.

For Student Two, errors <u>decreased</u> slightly after the time-only contingency was initiated. This is partly explained by the high error rate of the student in the first twelve sections of the program, going as high as seventeen and nine errors on sections eleven and twelve respectively. Omitting the poor performance on the beginning sections of the program for this student, we find that the time taken did drop after the time-only contingency was established, though only very slightly.

Student Three, as with the first, showed error increases after the time-only contingency was established. This student's performance showed an average of 4.17 errors in the first twelve sections, followed by a sharp decrease to 3.00 errors in the next thirteen sections. From sections twenty-six to thirty-four, however, the error rate rose to 3.79, as predicted.

All students showed time decreases of rather large magnitude following establishment of time-only contingencies. It is believed that these time decreases were not part of a possible "normal" decrease in time taken as a result of the students becoming more familiar with the material and the program.

In the case of one of the three students the drop following initiation of time-only contingencies was larger than the difference between time taken on the first twelve and the next thirteen sections. This student averaged 51.00 minutes on the first twelve sections, and 41.70 minutes on the next thirteen sections, a difference between the two subsections of 9.30 minutes.

However, during the time-only sections, the time averaged 29.55 minutes for this student, a difference of 12.15 minutes between these sections and the average time taken on sections thirteen to twenty-five, completed under both time and error contingencies.

Another of the students experienced a drop in time taken under time-only contingencies which was only slightly less than that experienced when comparing times on the first twelve sections with times on the next thirteen sections.

In this case, time on the first twelve sections averaged 44.08 minutes, and on the next thirteen sections 33.69 minutes, for a difference of 10.39 minutes. He averaged 24.08 minutes on the next 25 sections, the difference being 9.59 minutes.

The main evidence for the decrease in time occuring under time-only contingencies not being a "natural" phenomenon, but instead due to the effects of the reinforcement contingencies, comes from the third student. This student averaged 36.83 minutes on the first twelve sections, but his time increased during the next thirteen sections to 37.54 minutes. Following establishment of time-only contingencies, the time taken showed a dramatic decrease to 28.55 minutes.

It appears quite probable that the time-only contingency affected time taken and errors strongly, more so in the case of errors made in the program,

which had shown a decreasing trend prior to the establishment of the time-only contingency.

Of additional interest is the fact that, in the single section Student One completed during the third part of the Level 5 program (error-only contingency), an increase in time taken was found, to be exact, thirty-two minutes, while four errors were made.

It is true that performance on one section does not establish a trend. However, it should be mentioned that the time of thirty-two minutes made on the section completed under error-only contingencies was higher than that made in the program for the previous twenty-two sections! The thirty-two minute score, made on section fifty-one of the Level 5 program, was higher than any other time score going back to section twenty-nine, where the student took forty minutes to finish working on that section.

CHAPTER VIII

SUMMARY

One of the major aspects in testing the English program described in this report, was the comprehensive set of objective tests administered to the students. These tests, seven in number were administered to the students at two points: after they had completed Level 3, and following completion of Level 4.

In addition to an interest in the test scores per se, of equal interest is the effect on the scores of allowing the students to work on the tests to completion, following the Level 4 program.

Items of interest and importance were uncovered as a result of these procedures. Among these were the following:

(1) The myth that an intelligence test is necessarily a test of a person's ability to solve problems was somewhat tarnished. Results showed that the stringent time requirements of the Otis Group Intelligence Test put our culturally-disadvantaged students at a further disadvantage. The majority of questions answered were answered correctly, but many questions were missed because of lack of time. When allowed to work to completion, students got a high percentage of questions correct almost as high a percentage correct as they scored on the questions answered prior to time being called. Thus, most questions were not missed because they were too "hard", but simply because the students could not read fast enough to complete many questions. In some respects The Otis Group Intelligence Test appears to be a measure of reading speed rather than problem-solving ability.

(2) Allowing students to work to completion on other tests, such as the Metropolitan Reading Achievement Test, resulted not only in some very large score increases, but several score reversals, with regard to the supremacy of the student's performance in for example, , Word Knowledge and Reading. Thus, on the second administration of the Metropolitan Reading Achievement Test, the average scores for the students, within the time alloted, were a 8.17 grade level for Word Knowledge and a 5.36 grade level for Reading. Allowing students to work to completion on this test resulted in a reversal of the relative supremacy of the two scores in the test. The Word Knowledge score rose to an 8.77 grade level, while the Reading score increased greatly to a 9.00 grade level. As with the Otis testing, lack of time severly hampered these students, who generally possessed poor reading habits. Just as the Otis tests put our students at a great disadvantage due to the stringent time allowances, so did the other objective tests result in much poorer performances than would have been the case if the tests had stressed accuracy rather than speed.

(3) A perfect correlation was found between shifts in the students' relative ranking with regard to speed on the Level 3 and 4 programs and their relative ranks on Forms A and B of the Otis Group Intelligence Test. In the one case where Otis ranking stayed the same from the first test to retesting, the speed ranking remained stable from Level 3 to Level 4. In the two cases where Otis ranking went up, the speed ranking also went up. This correlation did not hold when comparing accuracy rankings on Levels 3 and 4 to rank on Otis Forms A and B. This information agrees with a previous statement that speed

is the critical factor in performance on the Otis test. This is probably true for the other objective tests as well. A Spearman Rank-Order Correlation Coefficient of (+) .875 was found to exist when correlating speed ranking on Levels 3 and 4 with Otis Group Intelligence Test ranking on Forms A and B (students working with time limit).

(4) When allowed to go to completion on Form A of the Otis Test, students showed an increase of 20.67 IQ points, on the average, over their IQ score on the same test wher required to work within the time limits. The resulting average IQ score of 106.00 put the students in the top 31.7% of persons, nineteen years of age and over, listed in the Otis scoring manual. Within the time limits, on Form A the average percentile rank was 11.33, placing the students in the lower 12% of all persons, nineteen years of age and over, as listed in the scoring manual.

(5) There was a rather wide dissimilarity between scores on the various tests administered. Some tests showed very high scores, such as the Metropolitan Achievement Test, where the students averaged, on Form Bm (time limit) 8.10 grade levels, and on Form Am (no time limit), where the students averaged 8.89 grade levels. Other tests showed low scores, such as the Ohio Scholarship Tests, Techniques in Reading Comprehension for Junior High School, where the students averaged, on retesting (time limit), in the 2nd percentile for a grade nine baseline. These results underlined the necessity of administering a battery of tests to the students. Results of a single test may differ widely from the results of another test, which supposedly is measuring similar abilities.

(6) For every student, there was an inverse relationship between the amount of speed increase and the amount of accuracy increase when comparing

Level 4 performance with Level 5 performance. Speed increases can easily lead to accuracy decreases, while accuracy increases often come at the expense of speed.

(7) Use of the equilibrium value in the Level 5 program appeared to be effective, though the effect seemed greater on time than on errors. Analysis of the data showed clearly that the use of so-called "adjusting schedules of reinforcement" was an effective device to control the student's rate of errors and rate of speed throughout the program. Where contingencies were changed, as in the second part of the program, performance on both measures changed to conform to the new contingencies.

Future Use of Adjusting Schedules.

Classroom application of the adjusting schedule or equilibrium value technique would seem to offer immense possibilities for improving student performance. Equilibrium values for time, errors, and other parameters, could be established at the beginning of a course for all students. The basis for establishment of such values could be: (1) an arbitrary assignment of the same values to all students, or (2) the assignment of values dependent on the student's previous performance in more elementary levels of the same subject matter or in other subjects in the school curriculum. In this manner students could work at their own rate of speed, subject only to the contingencies imposed by the particular equilibrium values in effect. Instead of assigning grades to the student's performance, another method may be put into operation. Equilibrium values may be substituted for grades. For example, let us suppose a student is taking a course in English grammar. Normally

the student must proceed through the semester by working on the same material as all the other students, taking the same tests on the same dates and on the same material. The usual criterion for allowing a student to move on to the next level course is that of obtaining a mark of at least a "C", or 70%, or whatever, on an objective test. All students failing to make a "C" or 70% must take the entire course over again. Likewise, students who score, for example, 90% or 99%, must start the next course at the same level as did those students who could only get 70% or 75% of the test material correct. Use of equilibrium values instead of grades would result in the student not only competing against others, but against himself as well. Points, counting toward such things as a field trip or movie, could be given every time the student met the equilibrium value or values in effect on any particular section of the course he was working on. In addition, the students own equilibrium values could be used to determine when he was ready to move on to more advanced material, or the next level course. Therefore, students would be pushed to raise their equilibrium values, since by meeting the contingency requirements they would gain points counting towards various reinforcing activities. Similarly, they would be pushed to raise their equilibrium values to the point where they could move on to new material. Since the equilibrium values would fluctuate over time, computers could be used to determine when individual students were ready to switch over to a new, more advanced, learning program. The computers could determine average equilibrium values and absolute increases in equilibrium values. In addition, decreases in average and absolute values could be computed so that, if necessary, the student could be directed to material more suited to his present ability. Equilibrium values for speed could be played off against values for errors

to slow down or speed up the student's progress through the course if the teacher suspected that the student was working too fast or too slowly. For example, if the teacher suspected that a student was working too fast and therefore making too many errors, the time contingency could be eased or eliminated entirely. In this case the only reinforcement available to the student would be gained through meeting the error contingency, thus probably resulting in a decrease in errors made and increase in time taken. The possibilities for such a system are infinite.

In conclusion, it appears that the experimental use of devices of programmed instruction and behavioral psychology in the field of education offers a veritable goldmine to those interested. The use of adjusting schedules and equilibrium values is only one of the devices which may contribute to improvement of present educational methodology.

BIBLIOGRAPHY

Berman, Mark L. The Development of Proficiency in the Use of the English Language in Groups of non Western Indigenous Peoples Through Programmed Instruction. V.I.B. Project, Paper No. 1., Tempe, Arizona: Department of Anthropology, Arizona State University, May 1964.

_____. Some Considerations in the Education of Indigenous Groups in the Southwest. Professional Paper No. SP-2148. Santa Monica, California: System Development Corporation, 19 July 1965.

Frase, L. T. The Effect of Social Reinforcers in a Programmed Learning Task. Tech. Rep. No. 11, NONR 1834-36. Urbana, Illinois: Training Research Laboratory, University of Illinois, 1963.

Parisi, Domenico. Social Reinforcement and Performance in Programmed Learning in Italy. Tech. Rep. No. 27, NONR 134-36. Urbana, Illinois: Training Research Laboratory, University of Illinois, 1965.

Stolurow, L. M., Frase, L. T., and Johnson, C. R. The Effect of Social Reinforcers in a Programmed Learning Task--Experiment II. Urbana, Illinois: Training Research Laboratory, University of Illinois, 1964.

Sullivan, M. W. Programmed English a Modern Grammar. New York: Macmillan Company, 1963.