THE EFFECTS OF TRAINING AND PRACTICE ON PURDUE PEGBOARD SCORES OF MODERATELY AND SEVERELY RETARDED ADULTS

by

Carol S. Parks

Approved by

William H. Knight
Chairman, Thesis Committee

Professor of Psychology

Bruce L. Bachelder
Director of Evaluation and Director of Habilitation Unit, Western Carolina Center

W. G. Snipes
Chairman, Department of Psychology

Richard H. Rupp
Dean of the Graduate School

The Effects of Training and Practice on Purdue Pegboard Scores of Moderately and Severely Retarded Adults

Carol Parks
Appalachian State University

Running Head: The Effects of Training and Practice

Abstract

An experiment involving twenty moderately and severely retarded men was implemented in order to determine the effects of individualized training and practice on Purdue Pegboard scores. The investigator administered the Purdue Pegboard in the standardized fashion to each of the ten subjects in the experimental group, trained each to a pre-determined criterion on the same test, re-administered the standardized test, and then administered each subtest of the Purdue repeatedly until the subject had attained the same score three consecutive times. Ten subjects in the control group were individually administered the Purdue Pegboard in the standardized fashion and, with no intervening training or practice, took the same test one week later. Individual comparisons revealed that the experimental group post-test was significantly higher than experimental group pre-test, for all subjects, $p < .0005$, while the mean post-test scores for the control group were slightly lower than the control group pre-test. Matched t's showed the experimental group scores after practice to be significantly higher than the second standardized scores for all subtests, $p < .005$. Results of correlations gave reason to believe that test reliability increases with practice, for the retarded. The discussion of the experiment centered on the practical use of this procedure in sheltered workshops, the importance of being able to define exactly what a test is

testing for, emphasis on allowing for acquisition of skills on a given task prior to measuring production on the same task, elevating the level of expectancy professionals have for the retarded, and making the overall evaluation period more meaningful for the client.

The Effects of Training and Practice on Purdue Pegboard Scores of Moderately and Severely Retarded Adults

Vocational Evaluation is often the retarded individual's first chance in the world of work. He is referred to a vocational evaluator whose responsibility is to assess the client's work behavior and to predict his work potential through a variety of techniques and procedures (Nadolsky, 1971). The question logically comes to mind as to whether or not those techniques and procedures genuinely fulfill the intense need of that first chance. There is a growing number of evaluators who are beginning to build more complete evaluation programs; however, many continue to use instruments which, by now, have become sacred cows (Rusalem, 1972). As presently used, the instruments tend to yield such low scores as to suggest little or no work ability; they "promote a reliance on screening out individuals who are difficult to train instead of developing training procedures with sufficient power to meet the needs of all trainees (Gold, Note 1)." Through previous diagnostic

---

The investigator gratefully acknowledges the guidance of Marc W. Gold, Ph. D., Associate Research Professor, University of Illinois at Urbana-Champaign; and the cooperation of the Developmental Services Center, Champaign, Illinois, and Western Carolina Center, Morganton, N. C.

evaluations, it has usually already been established, before the individual walks through the door, that he is "subnormal;" why then do we put him through another series of standardized tests, which most often only "prove" his preconceived handicap?

Perhaps a brief review of the present evaluation process, its aims, assumptions, and procedures, is appropriate, in order to substantiate the criticisms which follow. There is a great number of tests and measurements available today. The literature abounds with descriptions of these instruments and statistical significance of same; however, descriptions of their practical significance are difficult to find. For a comprehensive review of this literature, the reader is referred to Gold (1973); a very brief review is presented here. Some writers have considered motor performance to be one of the most predictive variables to date, correlating highly with work competence (Taylor, 1964; Windle, 1960). Indeed, many manual dexterity tests have been used with the retarded (e.g., Ferguson, 1958; Patterson, 1964). Some of these dexterity tests are: Pennsylvania Bi-Manual, Purdue Pegboard, Minnesota Rate of Manipulation, Crawford Small Parts, O'Connor Finger Dexterity, O'Connor Tweezer Dexterity, Hand-Steadiness, and Lincoln-Oseretsky (Buros, 1974). Using twenty-five retardates, with IQs ranging from 30 to 50, Tobias and Gorelick (1960) found a correlation of .54 ($p < .1$) between Purdue Pegboard scores and hourly average piece rate at disassembling screws. Assuming no training, they concluded that this test was, therefore, a useful instrument for predicting productivity on a simple sheltered workshop task.

Other currently used evaluation techniques include situational assessments, behavioral observations, work samples, and on-the-job-evaluations. Situational assessment is carried out through an effort to reproduce an actual work environment within the rehabilitation center and then to carefully record the clients' work behavior in that setting (ICD Rehabilitation and Research Center, 1974). Lynch (1973) suggested objective behavioral observation, if recorded accurately, as one of the most precise data-gathering techniques available. Another attempt to simulate an actual work setting, used extensively throughout the United States, is the work sample system, which involves standardizing and obtaining normative data on typical work tasks (Gold, 1973). Familiar batteries are: TOWER, JEVS, Singer Graflex, and Evaluation Tests. The basic concept of work samples provides an enlightened approach to vocational evaluation; this activity provides the client an understanding of the realities of work and an awareness of his own strengths and weaknesses which will enable him to make a meaningful vocational choice (Nadolsky, 1974). On-the-job-evaluation is becoming quite popular. In this type of evaluation, evaluators assess their client's abilities on

actual jobs in the community (Allen and Shinnick, 1973; Genskow, 1973).

At first glance, one would naturally conclude that the procedures outlined above comprise a comprehensive and appropriate program of evaluation for the retarded. Lately, however, more and more criticisms have been voiced, and there are numerous reasons to re-examine the basic assumptions of vocational evaluation:

1. Most evaluation instruments currently used with the retarded were developed and validated with the normal population.

2. Test instructions are verbal, despite the fact that a key problem in retardation is poor verbal ability, and assessment of verbal ability is not the major intent of vocational evaluation. The verbal instructions in the Crawford Small Parts, for example, leave many retarded individuals staring blankly at the examiner, with use of such words as "tweezers," "grip," "right angle," "collars," "flange," "plate," and "threaded holes." Although demonstration is allowed, there are no specific non-verbal instruction techniques outlined.

3. Evaluation staff are biased by low expectations of what the retarded can do, as a result of traditional testing techniques (Karan and Gardner, 1973). It will not suffice, however, merely to raise expectations; "procedures must be developed and implemented to realize and challenge these new expectancies (Gold, 1972, p. 525)."

4. In the commonly used dexterity tests, there is no conceptualization of the important distinction between acquisition and production. Distinction must be made between: (a) the length of time and conditions necessary for an individual to learn a task, that is, to reach a predetermined criterion, and (b) the rate of his performance after he has learned the task. Gold (1973) stated that the effects of acquisition are not even considered for their possible importance, and suggested that if this separation was carefully implemented, the result would be highly reliable and descriptive data, and that training and evaluation could appropriately occur at the same time. This would, in effect, make the evaluation period more meaningful for both the client and the evaluator.

5. Initial test performance scores do not predict later efficiency (Parker and Fleishman, 1961). Wolfensberger (1967) questioned the eagerness of many to seize upon motor test scores as predictors, advising that initial performance on motor tasks is an invalid predictor of ability after training. Gold (Note 1) firmly concluded that initial performance is predictive neither of performance on the same task after training, nor of future job success.

Vocational evaluators, in light of these criticisms, are forced to question the validity of the initial test scores of their clients. To what extent do these initial scores, which often fall below the first percentile ranking,

predict an individual's ability to perform a given task? Gold (1972) trained sixty-four moderately and severely retarded individuals to assemble a fifteen-piece bicycle brake, an accomplishment which seems far beyond expectations of laymen and professionals. This data, coupled with extremely low initial dexterity test scores, raises some serious questions which can no longer be ignored. It is this inconsistency between initial test scores and ultimate performance that the present investigator addresses.

In an effort to develop and implement an alternate testing procedure, this investigator initiated a study to determine how training, on a test such as the Purdue Pegboard, would affect a retarded individual's performance on the same test after training. Furthermore, the effects of practice, in the form of repeated testing, were examined. It appears that training and practice should produce significant improvement in scores; but the investigator was unable to find empirical support reported on this presumption. It was the specific purpose of the present experiment to investigate the effects of individualized training and repeated testing on Purdue Pegboard scores of moderately and severely retarded adults.

Method

Subjects

The subjects were twenty retarded men enrolled in two sheltered workshops. Descriptive data for the experimental and control groups are presented in Table 1. Workshop directors were asked to select young adult clients of the same sex, with IQs less than 60, excluding from selection clients with severe physical or sensory handicaps. One subject with an IQ of 61 was selected because none other was available. The experimental group was first selected, and all procedures were completed. The investigator then realized the need for a control group, at which time ten more subjects of equivalent IQs and ages were selected. Two subjects were dropped from the experimental group during the first week of the experiment, one because of absenteeism, the second because of refusal to participate. Both were replaced, using the same selection criteria as previously stated.

Apparatus and Materials

A Purdue Pegboard test kit and a stopwatch were used.

Procedure

Standard administration. Each subject in both the experimental and control groups was seated individually at a table in a quiet, well-lighted room, with the experimenter seated beside him. The experimenter administered the Purdue Pegboard to each subject in the standardized

Table 1

Descriptive Characteristics of Subjects

Group	IQ			Age			Institutionalization (in months)			Workshop Enrollment (in months)		
	Range	Mean	S.D.	Range	Mean	S.D.	Range	Mean	S.D.	Range	Mean	S.D.
Exper. (N=10)	29-54	44.1	7.72	21-27	23.4	2.55	0-236	105	97.7	9-98	46.7	31.8
Control (N=10)	29-61	42.8	9.61	19-28	23.0	3.53	45-270	122.5	58.89	2-110	52.6	32.78

fashion, using the Three-Trial method. Great care was taken to see that the same objective instructions and demonstrations, as delineated in the testing manual, were given to each individual in both groups. A schematic diagram of the basic experimental procedure is presented in Table 2.

<u>Training</u>. Each subject in the experimental group was individually trained by the experimenter to perform the Purdue Pegboard tasks. The subject was seated directly across from the experimenter. The subject's preferred hand was determined by asking him whether he is right-handed or left-handed, then by asking him to demonstrate how he would perform three different activities: throwing a ball, brushing his teeth, and writing his name. The experimenter moved to the side of the subject to demonstrate the first subtest. Using the hand which is the preferred hand of the subject, the experimenter placed five pins in the preferred-hand column of the board, in the way which training will duplicate. Moving back to a position directly across from the subject, the experimenter grasped the subject's preferred hand and guided it to the appropriate cup containing twenty-five pins. The subject's fingers were manipulated to grasp the pin between the thumb, the index finger, and the middle finger. The hand was then lifted, still grasping the pin, and moved to the top hole of the appropriate column. The subject's index finger was placed on the end of the pin, closest to that finger, in order to pull the pin into a vertical position. The subject's hand was then moved to a

Table 2

A Flow Diagram of the Procedure

Group	Treatment									
		Training to Criterion					Practice to Stability			
Exper.	Stand. Admin.	Pre-ferred hand →	Non-Pre-ferred hand →	Both hands →	Assem-bly →	Stand. Admin. →	Pre-ferred hand →	Non-pre-ferred hand →	Both Hands →	Assem-bly
Control	Stand. Admin. →					Stand. Admin.				

resting position beside the hole, so that the heel of the hand was resting soundly on the board and the fourth and fifth fingers were curved inward. This was to correct unsteadiness which prevented the pin from easing into the hole. With the subject's index finger remaining at the tip of the pin, the experimenter manipulated the subject's fingers so as to guide the pin into the hole with the index finger, then to release the thumb and middle finger, and finally release the index finger and lift the hand to reach for another pin. Direct manipulation continued as needed, or when error occurred. Some subjects needed direct manipulation in order to attain good arm flexibility and eliminate a stiff elbow and awkward wrist movement. Minimal verbal explanation accompanied manipulation. The word "good" was said when the subject adhered to manipulation instruction and when the subject began to demonstrate the correct movements without assistance. Criterion for this subtest was twenty-five consecutive correctly inserted pins, without assistance from the experimenter. Accuracy, rather than time, was the key factor in training. The experimenter removed pins varyingly from the board during training, i.e., after 10 had been placed, after 12, 19, 15, etc., so that the subject would not inadvertently conclude that success could be achieved only by completing the entire board. Training took place during thirty-minute sessions twice a day until criterion was reached. When the subject

reached criterion, the experimenter placed her hand on the subject's hand and said "very good." All training procedures for the non-preferred hand subtest were like those used with the preferred hand subtest.

The training procedure for the Both hands subtest consisted of combining the skills already mastered in the two previous subtests. Conditions, again, were the same. The experimenter emphasized the simultaneous and parallel movement of the hands, arms, and fingers. In order to train this skill, the experimenter placed both hands around the wrists of the subject, so that the subject could manipulate his fingers himself, and moved the hands together from the cups to the board and back again until the subject could perform same without assistance. Criterion for this subtest was twenty-five consecutive pairs of pins correctly inserted without assistance from the experimenter.

Each experimental subject was then trained in the Assembly subtest. The experimenter first demonstrated five complete assemblies. The subject was instructed to place a pin in the top hole of the preferred hand column with his preferred hand, as he had previously learned. The experimenter then grasped the subject's non-preferred hand and guided it to the cup containing twenty washers. The subject's fingers were manipulated to grasp one washer between the thumb and the index finger. With the palm facing down, the hand was guided to the positioned pin and lowered so that the index finger rested on the tip of the pin through the hole in the washer, and the washer was allowed to drop over the pin. The experimenter then grasped the preferred hand of the subject and guided it to the cup containing twenty collars. The subject's fingers were manipulated to grasp one collar between the thumb, index finger, and middle finger, and to maneuver the collar into a vertical position. The hand was guided to the positioned pin and washer, and the collar was allowed to drop over the pin. The non-preferred hand was grasped and guided through the washer assembly again, thus completing one assembly. Manipulation continued as needed through repeated assemblies until the subject demonstrated correct assembly and alternating movement of hands in ten consecutive assemblies. The experimenter, with use of both hands, then proceeded to guide the subject's gross arm movement in a simultaneous, alternating movement while the subject manipulated his fingers, so that, as one part was being positioned on the board with one hand, the other hand was picking up the next part, and so on. Manipulation continued as needed until the subject demonstrated correct assembly, alternating movement of hands, and simultaneous movement of hands in ten consecutive assemblies. Assembly parts were removed varyingly from the board, as in the other subtests. Training proceeded in thirty-minute sessions twice a day until criterion was reached. When criterion was reached on the Assembly

subtest, the experimenter placed both hands on the subject's hands and praised him freely.

Standard administration. The experimenter administered the Purdue Pegboard to each subject of the experimental and control groups in the standardized fashion, using the Three-Trial method.

Practice to stability. The experimenter administered each subtest of the Purdue Pegboard repeatedly in the standardized fashion to each subject in the experimental group, until each achieved the same score on three consecutive trials, that is, reached stability. Stability was reached in one subtest before moving to the next. Practice was held in thirty-minute sessions twice a day, with a one or two minute break between trials.

Results

Training

A 2 X 2 (Training X Experimental/Control) within subjects analysis of variance (Winer, 1962, p. 307) was performed on the raw scores obtained from the Purdue Pegboard standardized testing procedure.

Composite scores. Composite scores were obtained by totaling scores from the four subtests. The mean composite scores are presented in Figure 1. The main effects of both Training and Experimental/Control and their interaction reached significance (Table 3). Individual comparisons by

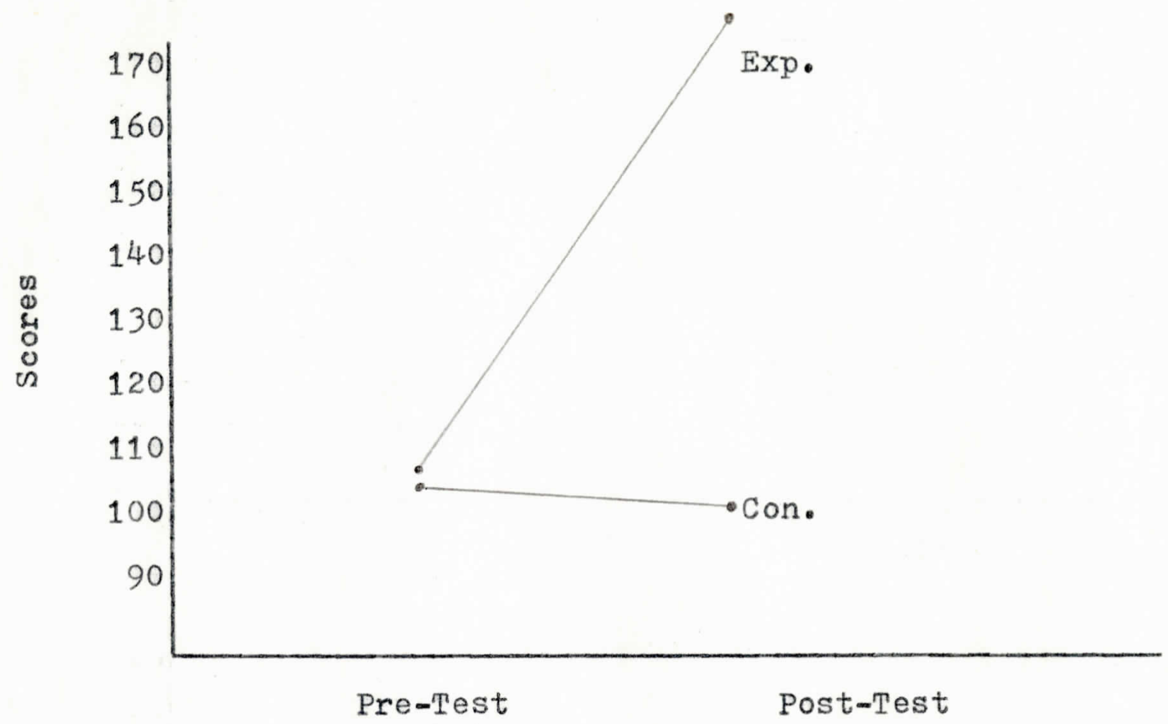

Figure 1. Composite pre- and post-test scores for experimental and control groups.

Table 3

Analysis of Variance on Composite Scores

Source	df	MS	F
Between subjects	19		
A (Group)	1	15,366.4	8.21*
Subjects within group	18	1,872.11	
Within subjects	20		
B (Training)	1	10,627.6	71.79**
AB	1	11,971.6	80.87**
B x Subjects within group	18	148.04	

*$p < .05$

**$p < .0005$

the planned comparisons method (Hays, 1963, p. 478) revealed that the experimental group post-test scores averaged higher than the experimental group pre-test, F (1,18) = 152.52, $p < .0005$. The mean post-test score for the control group was lower than the control group pre-test; however, this difference did not approach significance, F (1,18) = .14. The difference between the mean post-test scores of the experimental and control groups was significantly larger than the difference between the mean pre-test scores of the experimental and control groups, F (1,18) = 103.8, $p < .0005$.

<u>Preferred hand scores</u>. The mean preferred hand scores are presented in Figure 2. The main effects of both Training and Experimental/Control and their interaction reached significance (Table 4). Individual comparisons by the planned comparisons method (Hays, 1963, p. 478) revealed that the experimental group post-test averaged higher than the experimental group pre-test, F (1,18) = 64.93, $p < .0005$. The mean post-test score for the control group was lower than the control group pre-test; however, this difference did not approach significance, F (1,18) = .096. The difference between the mean post-test scores of the experimental and control groups was significantly larger than the difference between the mean pre-test scores of the experimental and control groups, F (1,18) = 57.49, $p < .0005$.

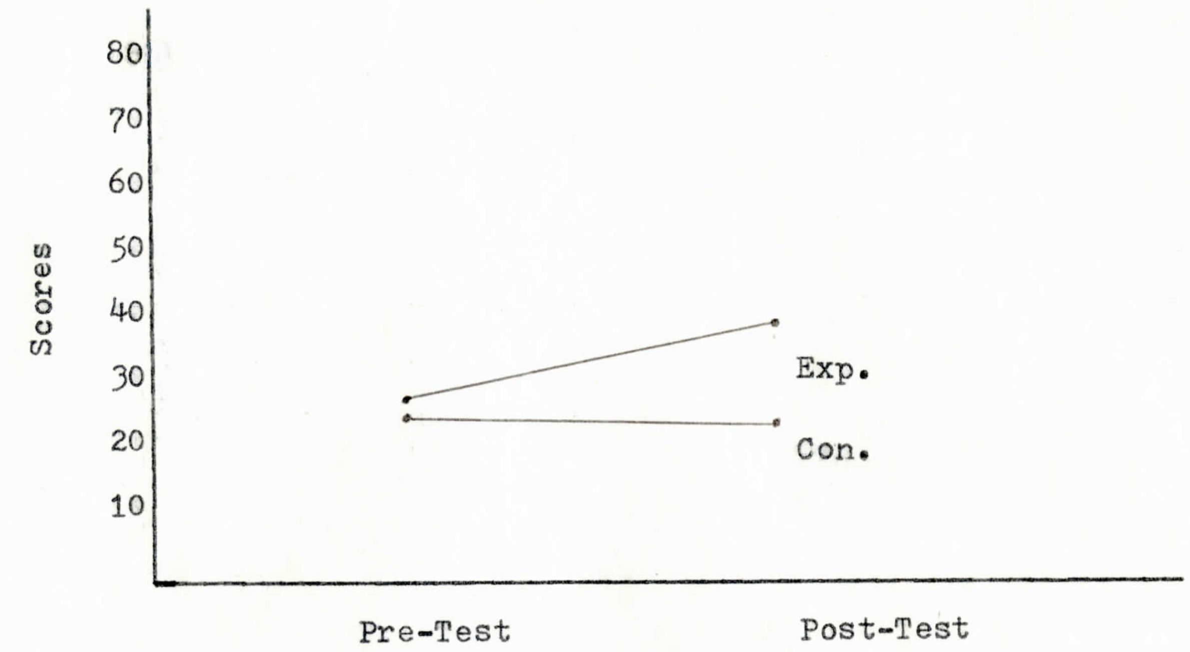

Figure 2. Preferred hand pre- and post-test scores for experimental and control groups.

Table 4

Analysis of Variance on

Preferred Hand Scores

Source	df	MS	F
Between Subjects	19		
A (Group)	1	748.22	11.43*
Subjects within Group	18	65.48	
Within Subjects	20		
B (Training)	1	390.62	30.02**
AB	1	455.62	35.01**
B x Subjects within Group	18	13.01	

*$p < .005$

**$p < .0005$

Non-preferred hand scores. The mean non-preferred hand scores are presented in Figure 3. The main effects of both Training and Experimental/Control and their interaction reached significance (Table 5). Individual comparisons by the planned comparisons method (Hays, 1963, p. 478) revealed that the experimental group post-test averaged higher than the experimental group pre-test, F (1,18) = 92.17, $p < .0005$. The mean post-test score for the control group was lower than the control group pre-test; however, this difference did not approach significance, F (1,18) = .04. The difference between the mean post-test scores of the experimental and control groups was significantly larger than the difference between the mean pre-test scores of the experimental and control groups, F (1,18) = 86.82, $p < .0005$.

Both hands scores. The mean both hands scores are presented in Figure 4. The main effects of both Training and Experimental/Control and their interaction reached significance (Table 6). Individual comparisons by the planned comparisons method (Hays, 1963, p. 478) revealed that the experimental group post-test averaged higher than the experimental group pre-test, F (1,18) = 59.36, $p < .0005$. The mean post-test score for the control group was lower than the control group pre-test; however, this difference did not approach significance, F (1,18) = .09. The difference between the mean post-test scores of the experimental and control groups was significantly larger than

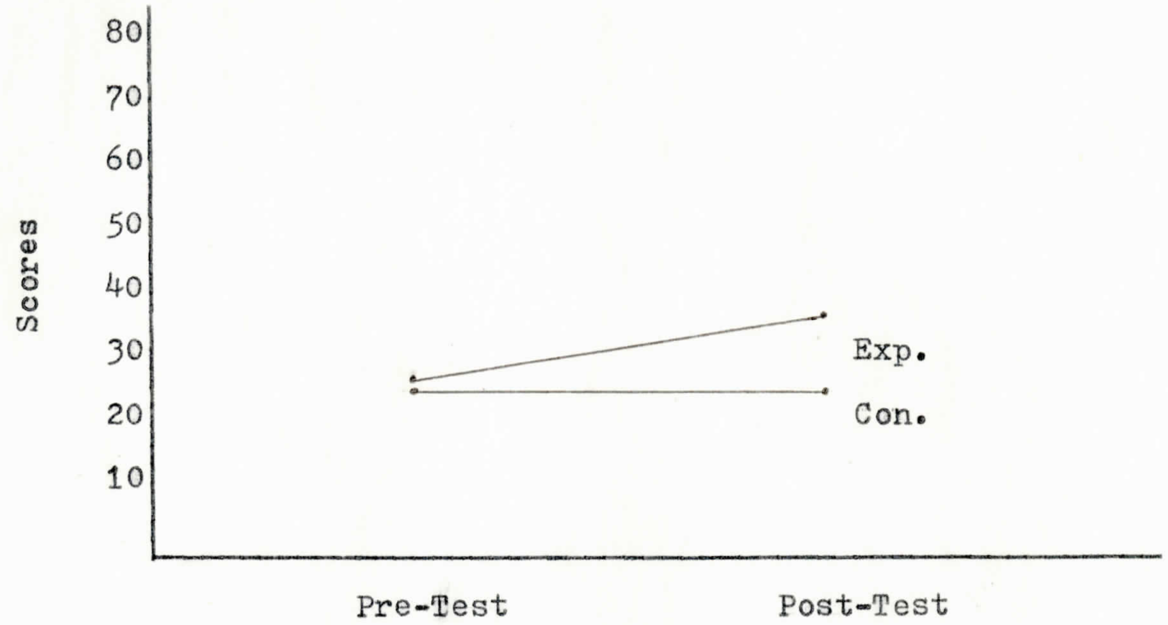

Figure 3. Non-preferred hand pre- and post-test scores for experimental and control groups.

Table 5

Analysis of Variance on Non-Preferred Hand Scores

Source	df	MS	F
Between Subjects	19		
A (Group)	1	490.	4.76*
Subjects within Group	18	103.02	
Within Subjects	20		
B (Training)	1	250.	44.3**
AB	1	270.4	47.91**
B x Subjects within Group	18	5.64	

*$p < .05$

**$p < .0005$

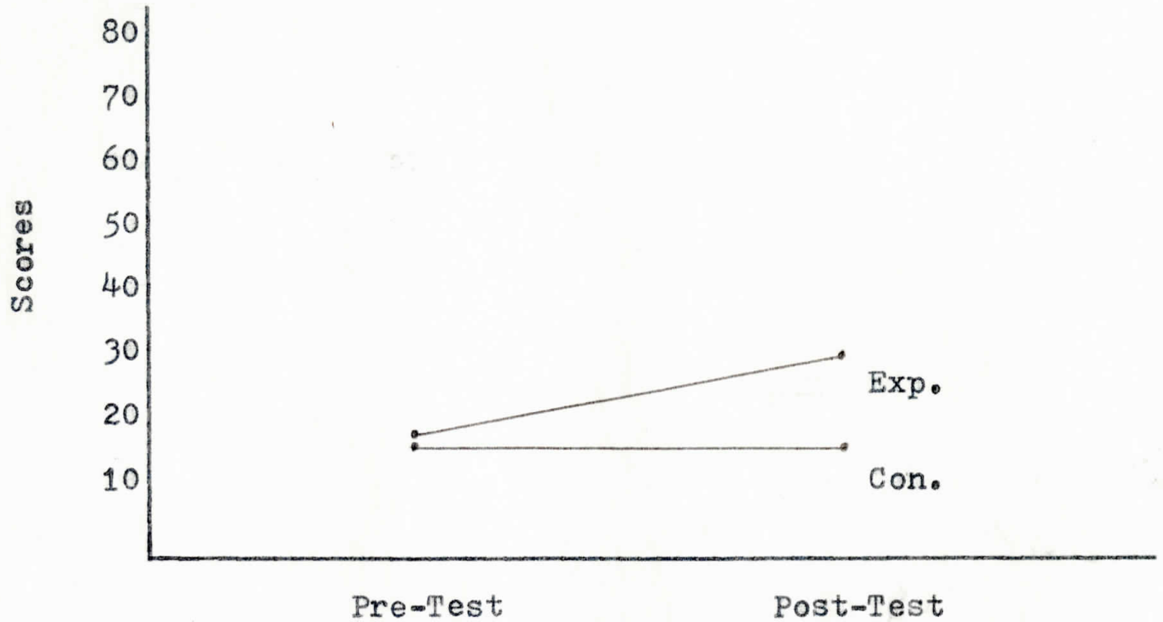

Figure 4. Both hands pre- and post-test scores for experimental and control groups.

Table 6

Analysis of Variance on

Both Hands Scores

Source	df	MS	F
Between Subjects	19		
A (Group)	1	366.03	5.42*
Subjects within Group	18	67.58	
Within Subjects	20		
B (Training)	1	235.23	27.38**
AB	1	275.63	32.08**
B x Subjects within Group	18	8.59	

*$p < .05$

**$p < .0005$

the difference between the mean pre-test scores of the experimental and control groups, F (1,18) = 42.6, $p < .0005$.

Assembly scores. The mean assembly scores are presented in Figure 5. The main effects of both Training and Experimental/Control and their interaction reached significance (Table 7). Individual comparisons by the planned comparisons method (Hays, 1963, p. 478) revealed that the experimental group post-test averaged higher than the experimental group pre-test, F (1,18) = 154.86, $p < .0005$. The mean post-test score for the control group was lower than the control group pre-test; however, this difference did not approach significance, F (1,18) = .11. The difference between the mean post-test scores of the experimental and control groups was significantly larger than the difference between the mean pre-test scores of the experimental and control groups, F (1,18) = 82.53, $p < .0005$.

Practice

A trend analysis (Winer, 1962, p. 133) was calculated on the practice series which followed the second standardized administration of the Purdue Pegboard test. The fact that each subject took the test repeatedly until his score stabilized, meant that each subject had a different number of test scores, and these scores could not easily be subjected to an analysis of variance. For this reason, the following dependent variable was developed to equalize the number of scores for each subject. Each subject's test scores on each subtest were combined algebraically into five scores. The number five was chosen because it was equal to the least

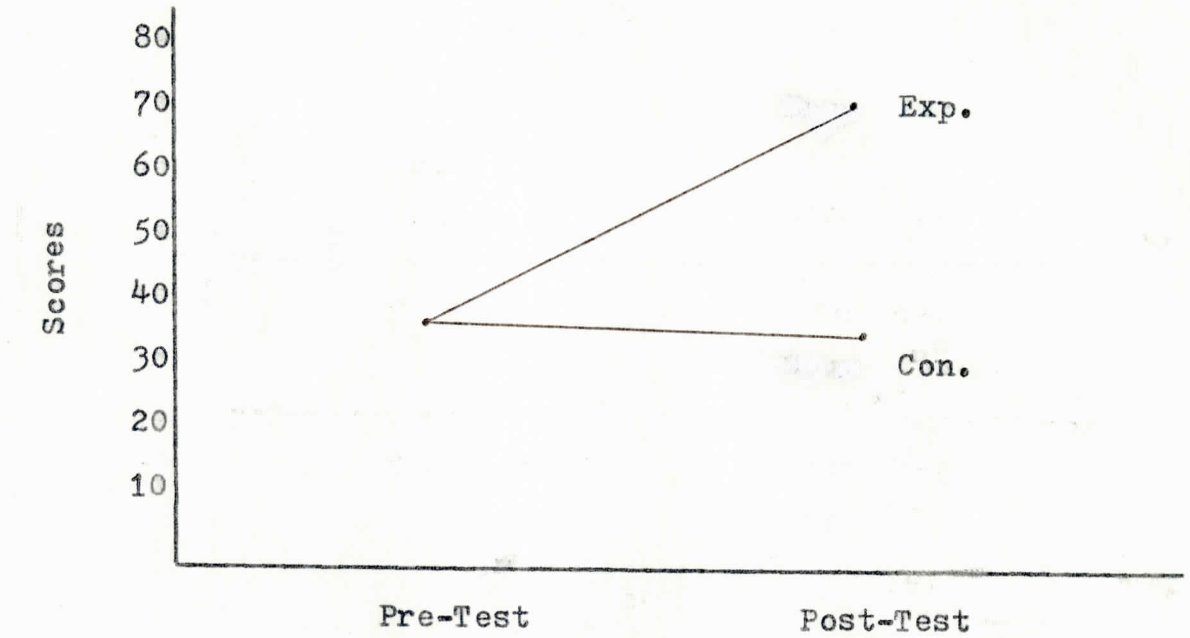

Figure 5. Assembly pre- and post-test scores for experimental and control groups.

Table 7

Analysis of Variance

on Assembly Scores

Source	df	MS	F
Between Subjects	19		
A (Group)	1	3,062.5	7.84*
Subjects within Group	18	390.73	
Within Subjects	20		
B (Training)	1	2,722.5	73.37**
AB	1	3,027.6	81.59**
B x Subjects within Group	18	37.11	

*$p < .025$

**$p < .0005$

number of trials taken by any of the subjects to achieve stability in test performance. It should be noted that this averaging procedure was conservative, in that the resulting five data points for each subject were equal to or less than the actual number of data points provided by that subject. Regardless of the number of test scores, all scores were reduced algebraically to five scores by a process of interpolation between actual test scores. First the number of actual test scores was divided by five. This yielded a number, d, which was used to determine the intervals used for interpolation. For example, a given subject had nine scores, N=9, e.g., 13, 14, 15, 14, 15, 15, 16, 16, 16; N÷5=d, or 9÷5=1.8. For each subject, d was multiplied by 1, 2, 3, 4, and 5, consecutively. These numbers indicated the interpolation intervals; d_1=1d, d_2=2d, d_3=3d, d_4=4d, and d_5=5d. Interpolation was based on the general formula: $\frac{Score_1 + .a(Score_2)}{1 + .a}$. Scores d_1, d_2, d_3, d_4, and d_5 were used to find the numbers for use in this formula, i.e., $Score_1$ = the score in the ordinal position, which is equal to the first digit of d_n; a = the second digit of d_n. Using the example of nine scores stated above, the following calculation is made to obtain the first data point:

$$\left.\begin{array}{lcl} d_1 & = & 1.8 \\ \text{First digit of } d_1 & = & 1 \\ a & = & .8 \\ Score_1\text{=first score} & = & 13 \\ Score_2\text{=second score} & = & 14 \end{array}\right\} \frac{13 + .8(14)}{1.8} = 13.444$$

the second data point:

$$\left.\begin{array}{ll} d_2 & = 3.6 \\ \text{First digit of } d_2 & = 3 \\ a & = .6 \\ \text{Score}_1\text{=third score} & = 15 \\ \text{Score}_2\text{=fourth score} & = 14 \end{array}\right\} \frac{15 + .6(14)}{1.6} = 14.625$$

the third data point:

$$\left.\begin{array}{ll} d_3 & = 5.4 \\ \text{First digit of } d_3 & = 5 \\ a & = .4 \\ \text{Score}_1\text{=fifth score} & = 15 \\ \text{Score}_2\text{=sixth score} & = 15 \end{array}\right\} \frac{15 + .4(15)}{1.4} = 15.0$$

the fourth data point:

$$\left.\begin{array}{ll} d_4 & = 7.2 \\ \text{First digit of } d_4 & = 7 \\ a & = .2 \\ \text{Score}_1\text{=seventh score} & = 16 \\ \text{Score}_2\text{=eighth score} & = 16 \end{array}\right\} \frac{16 + .2(16)}{1.2} = 16.0$$

the fifth data point:

$$\left.\begin{array}{ll} d_5 & = 9.0 \\ \text{First digit of } d_5 & = 9 \\ a & = .0 \\ \text{Score}_1\text{=ninth score} & = 16 \end{array}\right\} \frac{16 + .0(0)}{1.0} = 16.0$$

Thus, the five data points for this example are: 13.444, 14.625, 15, 16, and 16. This procedure was carried out for the scores of each subject on each subtest. Means of these

scores for the ten experimental subjects were calculated and are presented graphically in Figure 6. Trend analyses were calculated for all subtests and are presented in Tables 8, 9, 10, and 11.

<u>Preferred hand scores</u>. There was a significant linear trend in the scores; the quadratic trend component did not reach significance (Table 8). The final score after practice was significantly higher than the second standardized score, matched t (9) = 11.77, $p < .005$.

<u>Non-preferred hand scores</u>. There was a significant linear trend in the scores; the quadratic and cubic trend components did not reach significance (Table 9). The final score after practice was significantly higher than the second standardized score, matched t (9) = 8.37, $p < .005$.

<u>Both hands scores</u>. There was a significant linear trend in the scores; the quadratic trend component did not reach significance (Table 10). The final score after practice was significantly higher than the second standardized score, matched t (9) = 10.86, $p < .005$.

<u>Assembly scores</u>. There was a significant linear trend in the scores; the quadratic and cubic trend components did not reach significance (Table 11). The final score after practice was significantly higher than the second standardized score, matched t (9) = 8.57, $p < .005$.

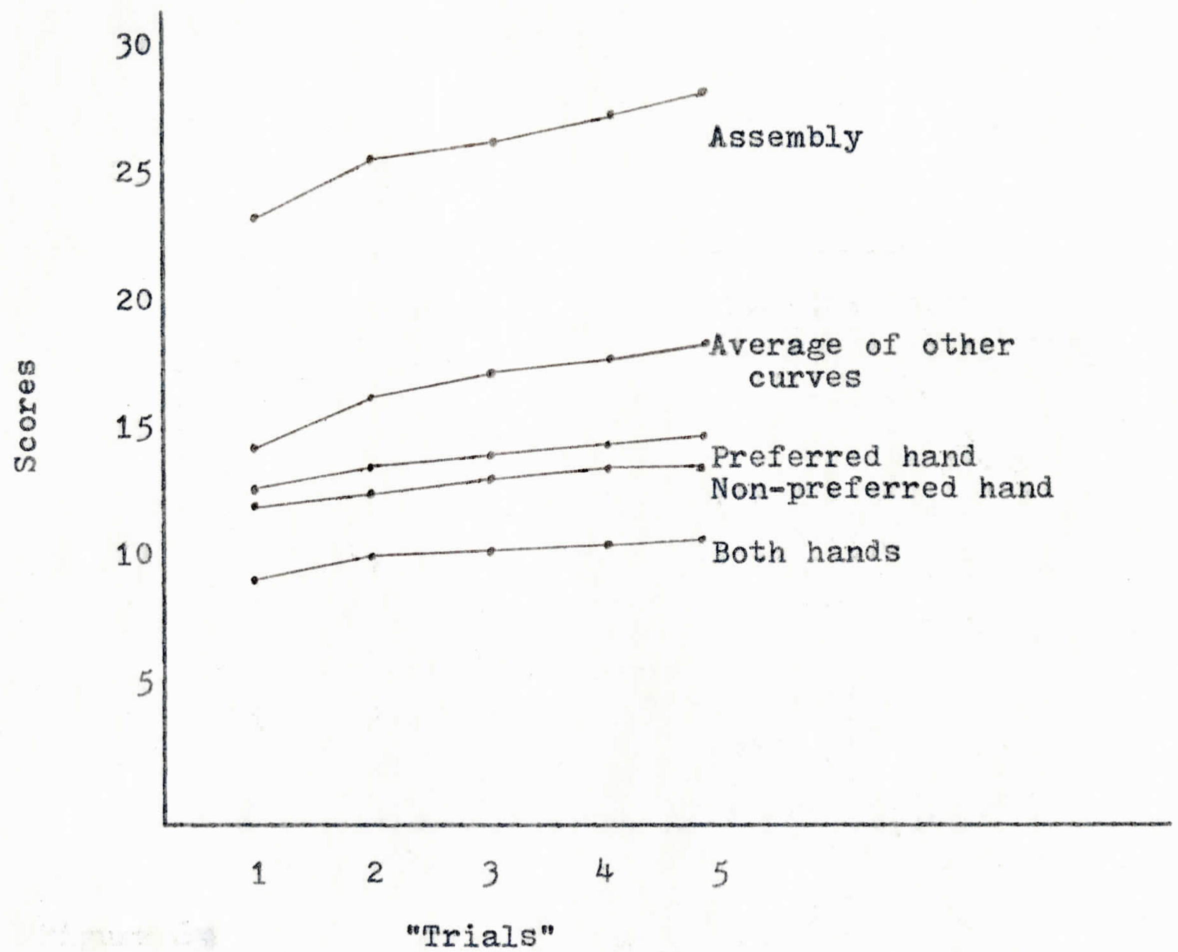

Figure 6. Means of interpolated practice scores for each subtest.

Table 8

Trend Analysis on Preferred Hand Scores

Source	*df*	*MS*	*F*
Between Subjects	9		
Within Subjects	40		
Trials	4	12.39	27.34*
Linear	1	47.9	105.74*
Quadratic	1	1.61	3.55
Residual	36	.453	

*$p < .0005$

Table 9

Trend Analysis on Non-Preferred Hand Scores

Source	*df*	*MS*	*F*
Between Subjects	9		
Within Subjects	40		
Trials	4	6.02	23.15*
Linear	1	22.53	86.64*
Quadratic	1	.48	1.86
Cubic	1	.27	1.03
Residual	36	.26	

$^*p < .0005$

Table 10

Trend Analysis on Both Hands Scores

Source	*df*	*MS*	*F*
Between Subjects	9		
Within Subjects	40		
Trials	4	6.58	15.21*
Linear	1	25.58	59.08*
Quadratic	1	.21	.49
Residual	36	.43	

$^*p < .0005$

Table 11

Trend Analysis on Assembly Scores

Source	df	MS	F
Between Subjects	9		
Within Subjects	40		
Trials	4	34.99	22.53*
Linear	1	130.04	83.73*
Quadratic	1	1.81	1.16
Cubic	1	4.94	3.18
Residual	36	1.55	

*$p < .0005$

Correlations

A correlation matrix was calculated on the pre-test scores, post-test scores, and stabilized scores after practice. This matrix is presented in Tables 12, 13, 14, and 15.

Percentile Scores

Mean percentile rankings, as recorded on the Standard Purdue Pegboard Profile Sheets, are presented graphically for the experimental group in Figures 7, 8, 9, and 10. Mean percentile scores for the control group, both before and after training, were less than one percentile.

Table 12

Correlations for Preferred Hand Scores

	1	2	3
1		.59	.57
2			.93
3			

a_1 = Pre-test score

b_2 = Post-test score

c_3 = Stabilized score after practice

Table 13

Correlations for Non-Preferred Hand Scores

	1	2	3
1		.91	.83
2			.94
3			

a_1 = Pre-test score

b_2 = Post-test score

c_3 = Stabilized score after practice

Table 14

Correlations for Both Hands Scores

	1	2	3
1		.72	.74
2			.96
3			

a_1 = Pre-test score

b_2 = Post-test score

c_3 = Stabilized score after practice

Table 15

Correlations for Assembly Scores

	1	2	3
1		.78	.68
2			.96
3			

a_1 = Pre-test score

b_2 = Post-test score

c_3 = Stabilized score after practice

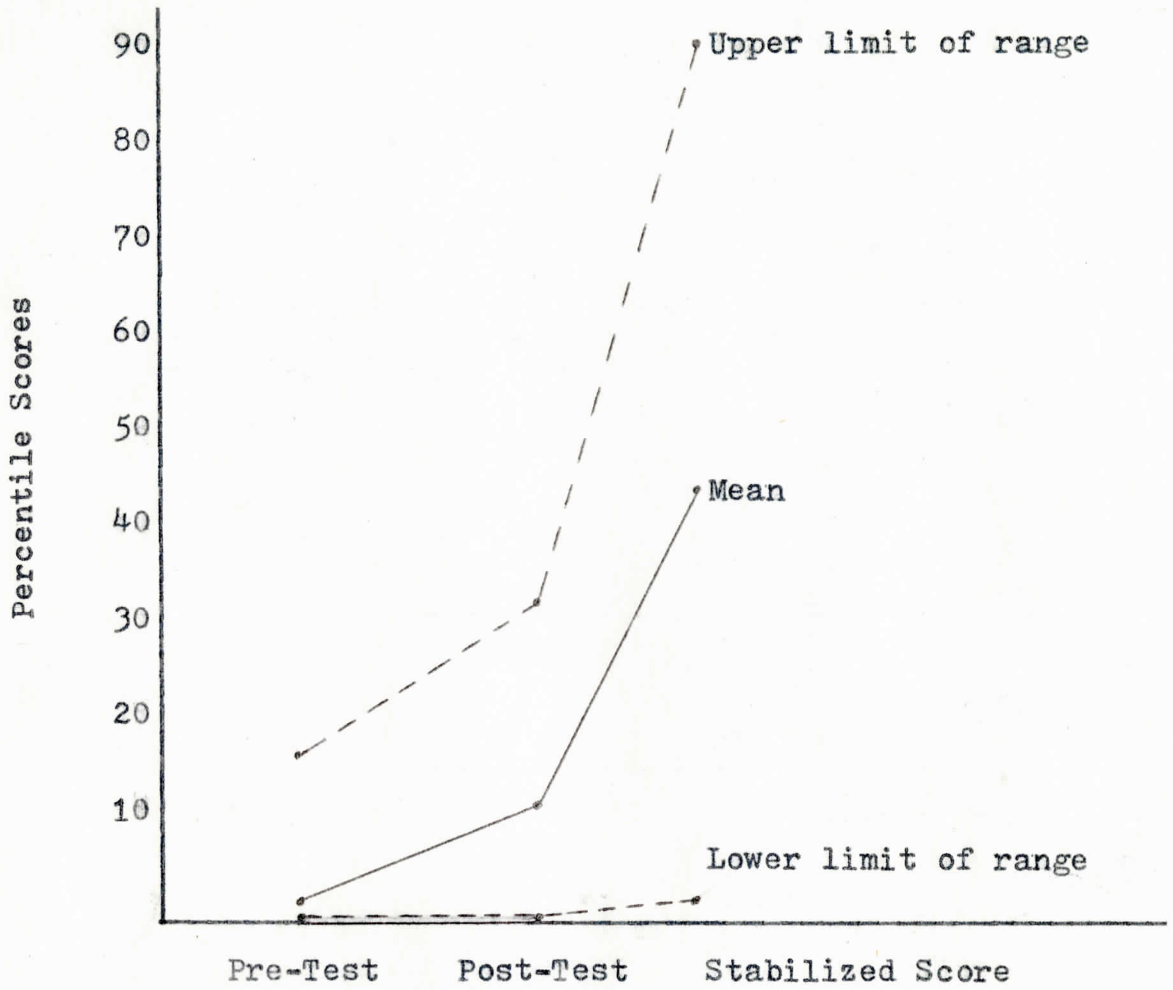

Figure 7. Mean percentile rankings of experimental group on the preferred hand subtest.

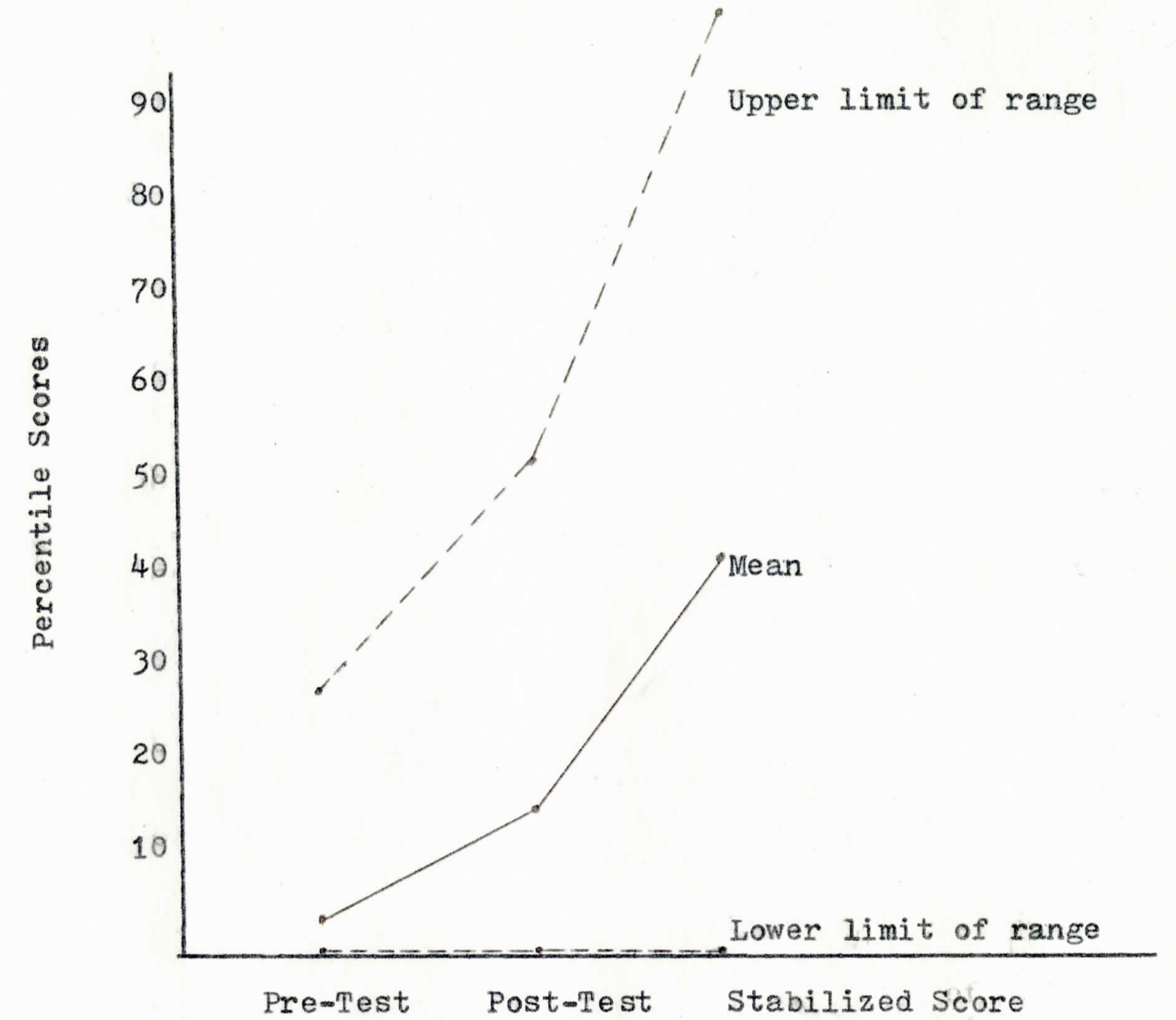

Figure 8. Mean percentile rankings on the experimental group on the non-preferred hand subtest.

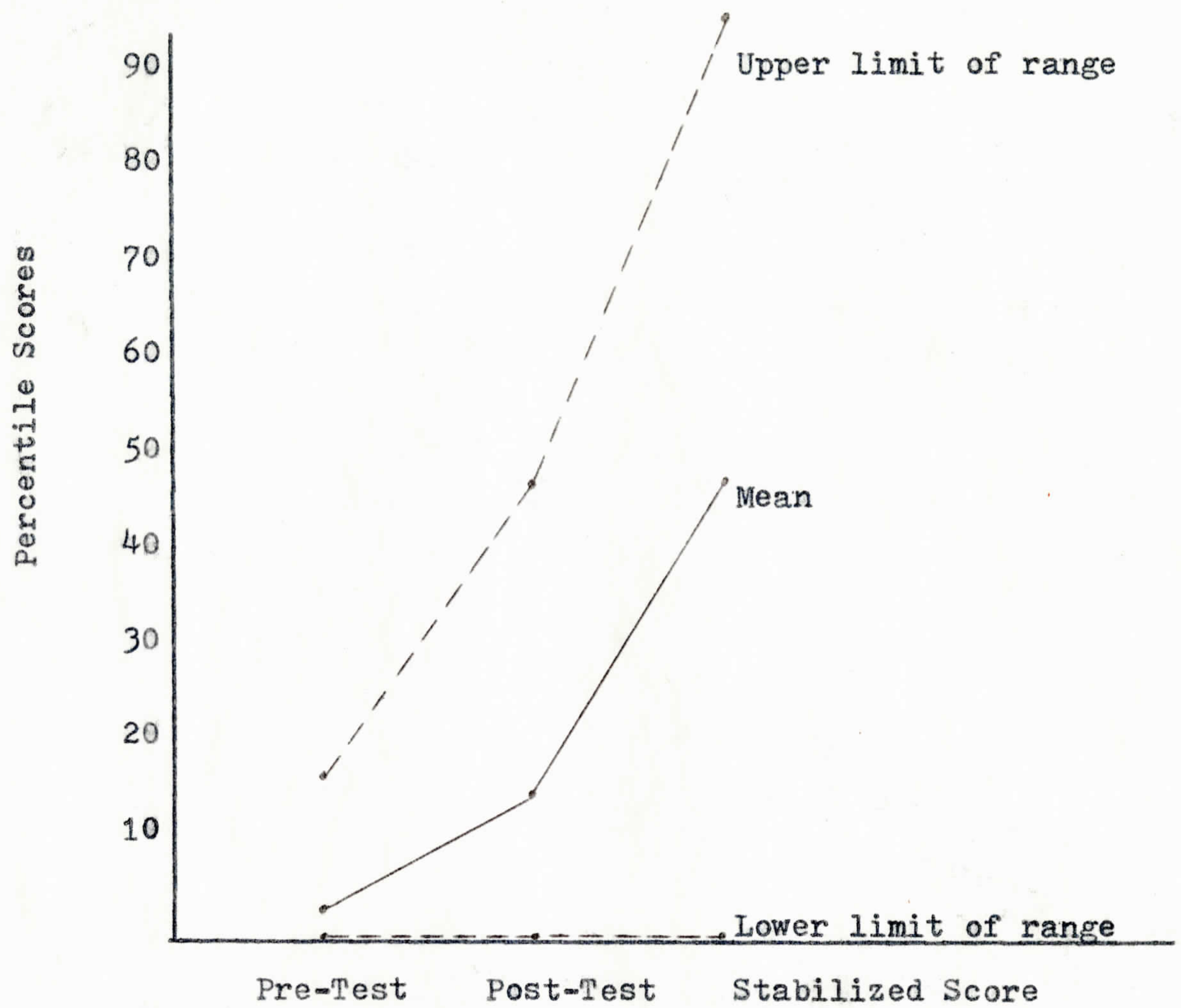

Figure 9. Mean percentile rankings for the experimental group on the both hands subtest.

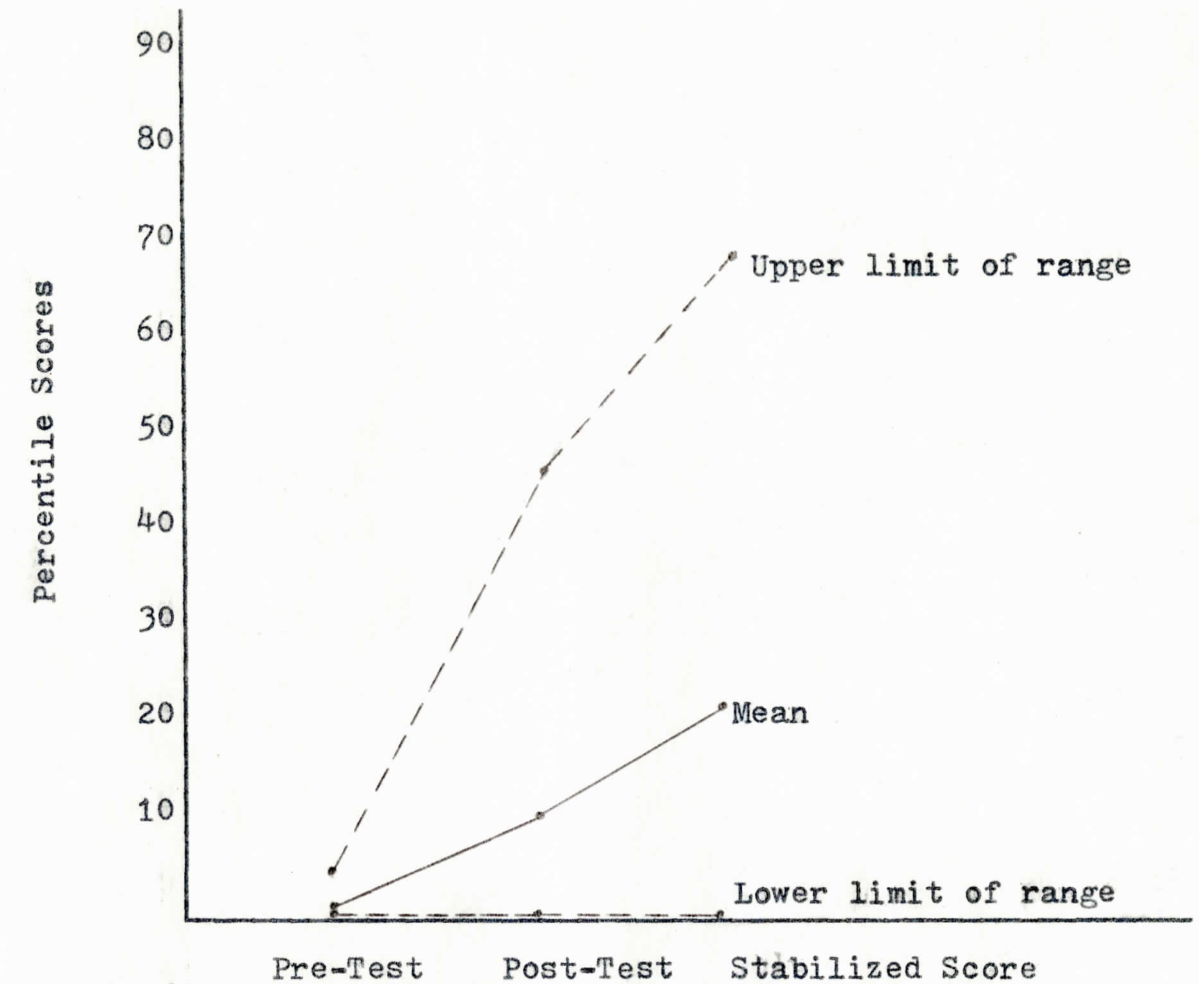

Figure10. Mean percentile rankings for the experimental group on the assembly subtest.

Discussion

As a result of this study, it was concluded that training of the component skills required on the Purdue Pegboard dexterity test results in much improved test scores. The experimental group improved significantly, while the control group actually declined somewhat on their second test. Performance of the experimental group continued to improve with practice following training, as revealed by a significant linear trend in the scores. These changes were not only significant, they were also important for their practical purposes, and have, therefore, raised some critical questions about a very accepted test.

In an ideal testing situation, one can readily define exactly the behavior which is being tested and the conditions under which it occurs. The Purdue Pegboard is designed to measure manipulative dexterity. In view of the large improvements made by the experimental group, an explanation must be sought, for it is not likely that the investigator took people without dexterity and gave them dexterity. The test must be testing something else, such as test wiseness, or past learning history; perhaps it is testing cognition (Gold, Note 2). In order to perform the tasks required on the Purdue Pegboard, one must use his cognitive functions; he must decide upon a cycle of movements he will employ. The more quickly he establishes a workable cycle and adheres to it, the better score he will

make on the test. The assumptions are that an individual will automatically establish a strategy, and that however fast he performs tells how his hands work; whatever measure is obtained is called production. However, one of the major differences between the normal and the retarded is the time required to learn a task (Wolfensberger, 1967). For those individuals who have difficulty learning, the evaluator needs to provide for acquisition, that is, to find out how the individual learns a skill, and how he reaches a criterion of performance in terms of quality. Then, actual dexterity and production can be examined. While the data cannot prove this, it seems reasonable that the improvements of the experimental group were due to acquisition of specific skills and strategies, and an increase in test-wiseness. For this reason, it can be tentatively concluded that the final scores approach validity and reliability as indicators of dexterity, and can be compared to scores of normal individuals. It is assumed that normal individuals, on the average, are not deficient in these variables; future research should study the same variables with normal persons. Correlational studies supported the fact that the retarded individual's scores became increasingly reliable as training and practice proceeded. It is very likely that the "stabilized" scores obtained represent a plateau on the learning curve, and that further practice would reveal even greater improvement.

Implications of these results suggest that evaluators working with the retarded should not rely upon initial test scores as valid predictors; they should develop training procedures for each test, which will result in more valid and reliable scores. If tests are used in a way not intended, that is, if training procedures are designed for them, test developers may say the standardization has been violated; but the way tests are currently used is violating the retarded individual, which is worse than violating the tests (Gold, Note 2). If the client learns the task after training, the evaluator knows he has the power to teach, and the client has the power to learn something the initial test said he did not know. Following such a procedure with actual tests, the evaluator should then gather acquisition and production data on a number of work tasks, of varying levels of difficulty. Gold and Scott (1971) described several procedures for training, including task analysis and sequencing. Evaluation of other factors, such as motivation, attention span, attitude toward supervisor, and general work habits could be systematically included. At the conclusion of the evaluation period, the evaluator will have an abundance of meaningful information about the client to pass along to the work adjustment coordinator, instructors, and rehabilitation counselors. Much more accurate recommendations will be possible, and therefore much more appropriate training programs and plans for

the client's vocational future can be designated. Furthermore, through such an evaluation program, the client himself will experience growth; he will discover for himself a previously untapped reserve, and perhaps even a brighter self-image. Elevated expectancies and an improved technology of testing and evaluation should promote the advancement of the moderately and severely retarded individuals in our communities.

Reference Notes

1. Gold, M. W. Vocational training. In Wortis, J., (Ed.): Mental retardation and developmental disabilities: An annual review, vol. 7. New York: Brunner Mazel, 1975. Book in press.

2. Gold, M. W. Vocational evaluation. Lecture delivered at the University of Illinois, Champaign-Urbana, April 21, 1975.

References

Allen, M., & Shinnick, M. D. Placement through a job trial approach to vocational evaluation. Vocational Evaluation and Work Adjustment Bulletin, 1973, 6, 29-33.

Buros, O. K. Tests in print II. Highland Park, New Jersey: The Gryphon Press, 1974.

Crawford, J. E. & Crawford, D. M. Crawford Small Parts Dexterity Test. New York: The Psychological Corporation, 1956.

Ferguson, R. G. Evaluating vocational aptitudes and characteristics of mentally retarded young adults in an industrial-agricultural workshop. American Journal of Mental Deficiency, 1958, 62, 787-791.

Genskow, J. K. Evaluation: A multipurpose proposition. Journal of Rehabilitation, 1973, 39, 22-25.

Gold, M. W. & Scott, K. G. Discrimination learning. In Stephens, W. B., (Ed.): Training the developmentally young. New York: The John Day Company, 1971.

Gold, M. W. Stimulus factors in skill training of the retarded on a complex assembly task: Acquisition, transfer and retention. American Journal of Mental Deficiency, 1972, 76, 517-526.

Gold, M. W. Factors affecting production by the retarded: Base rate. Mental Retardation, 1973, 11, 41-44. (a)

Gold, M. W. Research on the vocational habilitation of the retarded: The present, the future. In Ellis, N. R., (Ed.): International review of research in mental retardation, vol. 6. New York: Academic Press, 1973, 97-147. (b)

Hays, W. L. Statistics for psychologists. New York: Holt, Rinehart, and Winston, 1963.

ICD Rehabilitation and Research Center. Vocational evaluation: The key to occupational adjustment. Testing orientation and work evaluation in rehabilitation (TOWER). New York: ICD Rehabilitation and Research Center, 1974, 22-31.

Karan, O. C. & Gardner, W. I. Vocational rehabilitation practices: A behavioral analysis. Rehabilitation Literature, 1973, 34, 290-298.

Lynch, J. Work behavior observations and records. Vocational Evaluation and Work Adjustment Bulletin, 1973, 6, 15-20.

Nadolsky, J. M. Vocational evaluation theory in perspective. Rehabilitation Literature, 1971, 32, 226-231.

Nadolsky, J. M. The work sample in vocational evaluation: A consistent rationale. Vocational Evaluation and Work Adjustment Bulletin, 1974, 7, 2-5.

Parker, J. F., & Fleishman, E. A. Use of analytical information concerning task requirements to increase the effectiveness of skill training. Journal of Applied Psychology, 1961, 45, 295-302.

Patterson, C. H. Methods of assessing the vocational adjustment potential of the mentally handicapped. The Training School Bulletin, 1964, 61, 129-152.

Rusalem, H. A capsule research review. Rehabilitation Literature, 1972, 33(1), inside front cover.

Taylor, J. B. The structure of ability in the lower intellectual range. American Journal of Mental Deficiency, 1964, 68, 766-774.

Tiffin, J. Purdue Pegboard. Chicago: Science Research Associates, Inc., 1948.

Tobias, J. & Gorelick, J. The effectiveness of The Purdue Pegboard in evaluating work potential of retarded adults. The Training School Bulletin, 1960, 57, 94-104.

Windle, C. Prognosis of mental subnormals. American Journal of Mental Deficiency Monograph, 1962, 66(5, Pt. 2).

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

Wolfensberger, W. Vocational preparation and occupation. In: Baumeister, A. A., (Ed.): Mental retardation: Appraisal, education, and rehabilitation. Chicago: Aldine Publishing Co., 1967, 232-273.