

AN ANALYSIS OF RANDOM AND SYSTEMATIC CHANGES WITH PRACTICE*

EDWARD B. GREENE
UNIVERSITY OF MICHIGAN

Six motor tests and six nonverbal tests were administered four times to the same subjects. Subjective reports of the subjects are discussed, changes in mean scores and in variability and score correlations from trial to trial are surveyed, and factor analyses of results on the first and fourth trials are presented and compared. Implications of the findings with respect to correction for attenuation are pointed out.

The Problem

Many articles have been written concerning retest-correlations and split-half correlations which assume that a low coefficient signifies the presence of only chance variations, either in the individuals tested or in the methods of observing and recording their performances or in both. This assumption underlies the formulas for correction for attenuation and the Spearman-Brown split-half prediction formula. The correction for attenuation, urged by Spearman (1, 2) and many others, may in reality be eliminating or reducing important trends of a systematic sort. Although it is doubtless true that random errors and variations in performance will tend to lower self-correlations, the opposite, namely, that a low self-correlation indicates chance variations, is by no means true. Systematic errors of measurement and systematic changes in performance will also lower self-correlations. In the repetition of a test, it is almost inevitable that systematic changes in performance will be present. Likewise the first and last halves of many tests call out different ability patterns. Hence it is important in a careful analysis to know what sorts of changes are affecting the results. There are two approaches to this problem, one descriptive and the other statistical. In the analyses reported here both approaches were used because they supplement each other. The

* Thanks are due to Mr. Searles and his staff at the Henry Ford Trade School, who gave both suggestions and aid in completing this work; to Dr. L. L. Thurstone, who gave suggestions for certain aspects of the analysis; and to L. R. Tucker and Robert Blakey, who made the final factorial analyses and suggested various conclusions.

two prerequisites for a careful study of practice effects are generally agreed to be unambiguous measuring instruments, and a constantly motivated group which remains intact. These prerequisites were secured in part by administering four forms of the *Michigan Non-Verbal Series* (Greene, 1931), one a day, to 394 boys in training at a large mechanical trade school. The boys ranged in age from 14 years and 2 months to 15 years and 8 months, mean 14 years and 9 months. Most of them had finished the eighth grade in the Detroit area and they were in general selected from the upper half of the scholarship ratings.

The *Michigan Non-Verbal Series* has four forms which have been carefully equated for difficulty at two or three levels of complexity. The series includes twelve sub-tests which were always given in the following order:

1. *Aiming*: with pencil, left hand, at 4mm. circles for one minute.
2. *Aiming*: with pencil, right hand, at 4mm. circles for one minute.
3. *Aiming*: with pencil, left hand, at $1\frac{1}{2}$ mm. circles for one minute.
4. *Aiming*: with pencil, right hand, at $1\frac{1}{2}$ mm. circles for one minute.

(Each aiming test was preceded by a 20-second warming-up work period. A standard No. 2 pencil of standard sharpness was used. All subjects were required to proceed from left to right across a row of five circles, then proceed from left to right on the next row below, until five rows were completed.)

5. *Tapping*: with pencil, left hand, 10 seconds, best of three trials.
6. *Tapping*: with pencil, right hand, 10 seconds, best of three trials.
7. *Feature Discrimination, Easy*: to circle with a pencil the unique feature in a row of *three* small printed faces, otherwise identical, 2-minute period.
8. *Feature Discrimination, Medium*: to circle one unique feature in one of *five* small faces printed in a row, 3-minute period.
9. *Feature Discrimination, Hard*: to circle with a pencil the unique feature in one of *seven* small faces printed in a row, 5-minute period.
10. *Maze Solution, Easy*: to draw a pencil line through rows of small square printed mazes, each having five alleys per side, without crossing any printed lines, 2-minute period.

11. *Maze Solution, Medium*: to draw a pencil line through square printed mazes, with 20 alleys per side, without crossing any printed lines, 3-minute period.
12. *Maze Solution, Hard*: to draw a pencil line through rectangular mazes ten times as large in area as those in No. 10, without crossing any printed lines, 5-minute period.

Four trials were given using the standardized equivalent forms, A, B, C, and D of these tests. Only the scores of boys who completed every trial of each test were included. The test scores were the number of correct responses in a period of time which was thought to be long enough to give a good sample of performance but not long enough to cause fatigue.

The tests were given in small class rooms during regular school periods in both morning and afternoon. Each group was tested on four consecutive days at the same time of the day. The motivation and effort seemed high and fairly constant. The feature discrimination tests were rated as less interesting than the other types of tests by nearly 90% of the entire group. The effects of interest on the scores have not been evaluated.

Descriptive Analysis of Processes

In order to give a clear picture of processes involved in success in the test situation, a summary of reports by subjects and observers is given. These reports were obtained from: 1) interviews with thirty subjects who were chosen to represent superior, average, and inferior performance on each type of test; 2) reports of six persons who observed trials of these thirty subjects, and others as well; and 3) reports of ten University of Michigan students who took the entire series of tests and wrote subjective analyses of their experiences after each trial. These reports indicate changes with practice toward the elimination of unnecessary movements and of acts usually associated with emotional behavior, such as smiling, scowling, coughing, etc. Many references are found to greater calmness in the later test situations, although none of the subjects were regarded as unusually emotional. An attempt was made to secure good motivation by telling the students that fair samples of their ability were desired and that each would be given a report of his test results at the conclusion of the series.

On the initial aiming tests the dominant processes seemed to be a flexible coordination of forearm and finger muscles and a rhythmic grouping of circles into horizontal sets and pauses. The more rapid and accurate students used very small finger and wrist move-

ments which gradually and directly approached the circles, with a well-marked rhythm of four or five dots, then a pause. The slower and more inaccurate students used large forearm and shoulder movements instead of, or alternating with, the finger and wrist movements. They exhibited irregular rhythms with two or three dots between pauses. They made the pencil point approach the circle with a ballistic or hammering movement which was sometimes checked and redirected before reaching the paper. They made heavier marks on the paper and more slips after placing the pencil than the faster students. The final trials showed a marked tendency toward smaller, more direct, less forceful movements, more regular rhythms, larger groups of circles, and shorter pauses between groups. The right-handed persons showed the heavy, ballistic, irregular forearm movements with the left hand, and reported greater effort and fatigue in left-hand aiming. The smaller circles required considerably more precision of movement than the larger ones.

On the initial tapping tests ballistic wrist and forearm movement seemed universal. The subjects were instructed to let the wrist lie comfortably on the table, but this was difficult for many students, especially with the non-dominant hand, and for slow subjects with either hand. The non-dominant hands used longer strokes, and hit the paper harder and more irregularly than faster hands. The shortest strokes were raised approximately one-quarter inch above the paper, and the longest, two inches. With practice, both hands tended to use shorter, lighter, and more regular strokes. No regular rhythms could be observed with either hand, and no spatial patterns were found. Students were asked not to try to tap a pattern. Most of them succeeded in letting the taps fall in a random fashion over an area of approximately one and three quarters inches in diameter.

The main process involved in the initial trial of the easy feature discrimination was a visual comparison of pattern. This was not dependent upon fine visual acuity. Each pattern was printed in large black lines and differed very noticeably in only one of the following ways: position, curvature, and length of line, or size and position of the dots representing eyes. The most successful students first compared all three faces with respect to one feature, such as hair. If the hair in all three was alike, some other feature, such as mouth, was selected and compared. This selection, comparison, and rejection was continued until the unique feature was found. The student then circled it with a pencil and proceeded to the next problem. Success in this test depended upon rapid and systematic isolation of a feature for comparison and upon rapid comparison of the three faces. Poor

scorers tried to compare whole faces, or repeated the comparison of features after they had been found to be uniform. The commonest error was the circling of one of two similar features. This error was due, according to subjects' reports, to a failure to compare all three faces. Some of the poorer scores were made by persons who isolated an attribute, such as length, and made a comparison of all features on this basis, overlooking for the time being any other differences such as curvature, position, or size. Changes during practice were generally in the direction of more rapid selection and isolation, more methodical comparison, and of the inclusion of more than one attribute of a feature in a comparative act. The comparison order changed from a somewhat random one, to a left to right sequence for nearly all subjects. The act of circling the unique feature with a pencil became more rapid with practice and a little less precise.

The medium feature discrimination tests included the same processes as the easy test just described. The muscular acts seemed nearly the same for all levels. The main differences between levels of difficulty are all reported to be in comparison processes. The comparison of a row of five faces demanded a more extended span of comprehension and a more methodical comparison than that of three faces. Avoidance of the repetition of comparisons became much more difficult and important. A slight distraction or emotional blocking would often make it necessary for the student to begin a problem all over again. Practice effects were also noted in a calmer and more determined approach to the task, a greater familiarity with the possible patterns in the test.

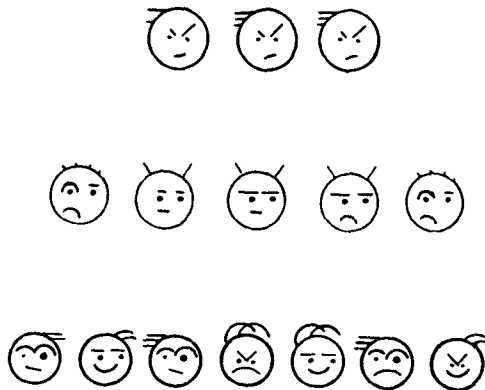


FIGURE 1
Samples from the Feature Discrimination Test

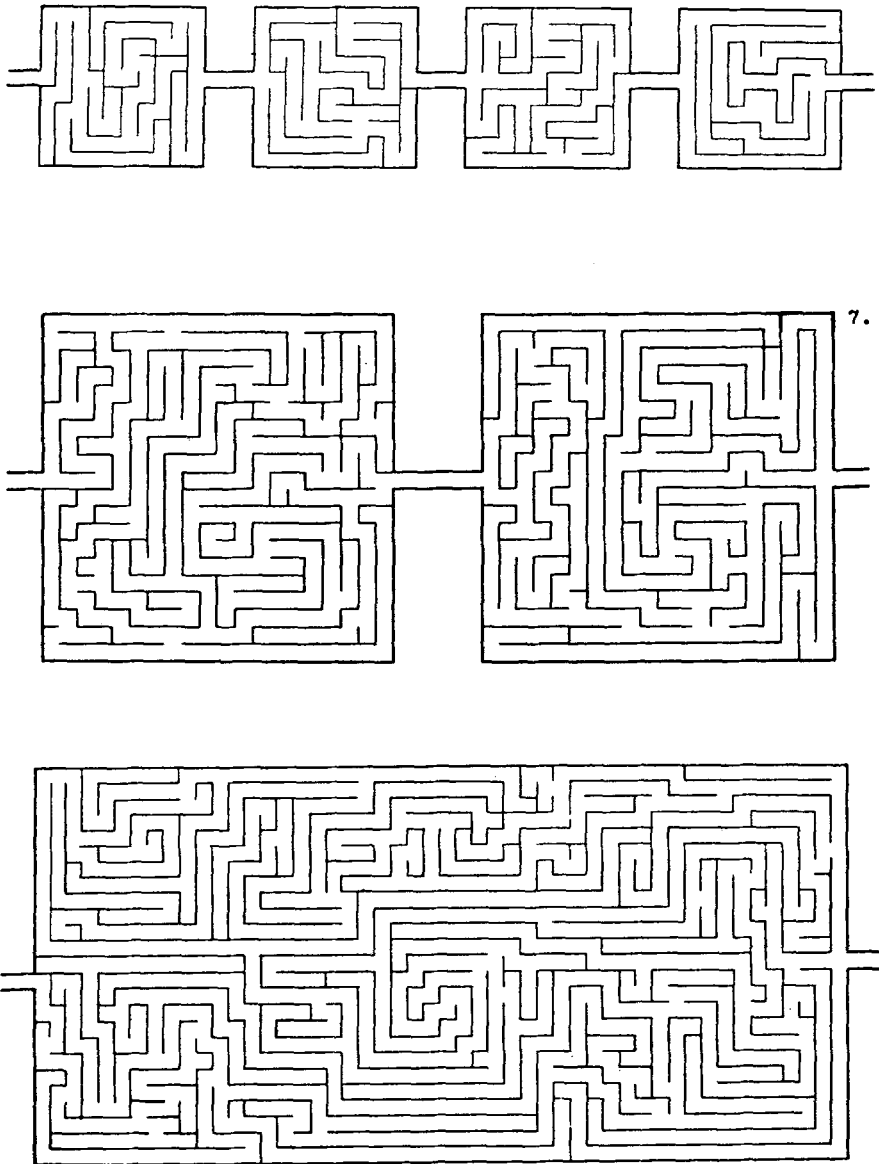


FIGURE 2
Samples from the Maze Test

The hard feature discrimination tests necessitated a comparison of a row of seven faces in which each feature always had three variations without showing a unique one. (Figure 1). The task then was to find a fourth variation which was different from any of the rest. Changes in skill seemed much like those noted above; the emotional elements and the concentration needed on each problem were much greater. A few students complained of fatigue and eye strain during the five-minute periods used.

The initial trial of the easy maze test seemed to be dominated by processes of simple visual comparison and of drawing lines. The comparisons of two pathways to see which led through to the next junction or to the end of the maze were not difficult for the more able subjects, since the blinds averaged 4 cm. in length. (Figure 2). Their test scores depended largely upon speed of drawing a line through alleys where the true pathway was perceived at a glance. Those making low scores may be divided roughly into two classes, the cautious, and those poor in comparison. The cautious persons generally made few errors and showed a more rapid improvement with practice than the others. Among those poor in comparison there was a considerable change with practice toward a longer comprehension span, fewer repetitions of entries into the same blind, more alertness in detecting openings and fewer emotional disturbances when no true path could be quickly found.

The medium maze test contained much more difficult comparisons, since the blinds averaged 8 cm. in length. Most of the students could not take the longest blinds in at a glance, but had to draw a pencil line some distance before discovering the dead end. Back tracing and repeated entries into the same blind were common at first, but were gradually reduced with practice.

The hard maze test differed from the medium in having greater length of blinds, averaging 12 cm., and greater complexity of pattern. With the larger area it was possible to make the true path follow spirals and to go long distances in the direction opposite from the goal, both of which were sources of error.

With practice, a skill was developed in avoiding the longer blinds, by means of visual explorations, and avoiding the repeated entries into blinds. Speed of drawing seemed to have little effect upon the score and it sometimes would lead to wrong decisions. The students were told to try to avoid blinds in order to get through the maze, but there was no penalty attached to errors other than the necessity of withdrawing from a blind. From these descriptive analyses, it seems fair to conclude that changes in performance from one trial to the

TABLE 1
 Four Trials of the Michigan Non-Verbal Series, Amounts Correct; Raw Score
 Means, Standard Deviations, and Coefficients of Variation

Form A—1st Trial				
Test	Mean		S.D.	C.V.
Aim Left, Large Circles	49.57		12.21	24.63
Aim Right, Large Circles	52.54		20.25	38.54
Aim Left, Small Circles	26.94		10.02	37.53
Aim Right, Small Circles	45.38		18.35	42.52
Tap Left	46.02		14.04	30.45
Tap Right	57.18		18.44	31.81
Feature, Easy	14.23		6.44	45.47
Feature, Medium	6.66		2.73	40.99
Feature, Hard	5.36		3.61	67.72
Maze, Easy	14.15		4.54	31.34
Maze, Medium	18.81		7.14	37.96
Maze, Hard	9.56		9.92	107.05
Form B—2nd Trial				
Test	Mean	% Improvement	S.D.	C.V.
Aim Left, Large Circles	51.81	4	13.8	26.63
Aim Right, Large Circles	71.26	36	16.55	23.27
Aim Left, Small Circles	31.10	15	8.68	28.05
Aim Right, Small Circles	48.35	6	12.56	26.10
Tap Left	51.21	11	8.46	20.55
Tap Right	59.94	5	4.32	7.33
Feature, Easy	19.58	38	5.88	29.72
Feature, Medium	9.40	42	3.77	40.11
Feature, Hard	6.35	18	2.11	48.57
Maze, Easy	17.98	26	5.56	31.06
Maze, Medium	18.31	-3	7.82	42.71
Maze, Hard	14.62	53	7.74	53.08
Form C—3rd Trial				
Test	Mean	% Improvement	S.D.	C.V.
Aim Left, Large Circles	60.07	21	16.10	26.81
Aim Right, Large Circles	72.46	38	18.0	24.84
Aim Left, Small Circles	32.86	22	10.98	33.20
Aim Right, Small Circles	50.25	11	12.12	24.01
Tap Left	53.91	17	9.06	16.72
Tap Right	63.02	11	8.28	13.26
Feature, Easy	21.23	50	6.02	28.33
Feature, Medium	12.58	99	4.68	37.25
Feature, Hard	9.47	76	3.90	41.17
Maze, Easy	17.92	26	5.84	32.41
Maze, Medium	24.05	28	8.79	36.55
Maze, Hard	20.28	112	10.62	52.28
Form D—4th Trial				
Test	Mean	% Improvement	S.D.	C.V.
Aim Left, Large Circles	60.42	22	15.25	25.24
Aim Right, Large Circles	77.35	46	19.72	25.48
Aim Left, Small Circles	33.06	22	10.38	30.64
Aim Right, Small Circles	51.9	14	13.24	24.43
Tap Left	55.9	21	10.62	19.38
Tap Right	72.44	26	12.40	17.37
Feature, Easy	22.73	59	6.34	27.93
Feature, Medium	14.14	113	4.28	30.27
Feature, Hard	8.66	61	3.62	48.01
Maze, Easy	19.28	35	5.50	28.37
Maze, Medium	25.63	38	10.53	41.08
Maze, Hard	20.30	114	11.13	50.02

next were to a large degree qualitative. Even in the simplest motor test, tapping, there was a change in the use of muscles holding and moving the pencil which can hardly be described as random, since there was a definite pattern of shift from large to small muscle groups. Evidence of changes of a random sort were principally small variations in the expenditure of energy in a given time.

Statistical Results

In evaluating the random and systematic variables in successive trials of a test, no single measure is sufficient. It is necessary to show how changes in self correlations are related to changes in raw scores, and to change in factor loadings. Table 1 shows improvements in raw score means, and standard deviations. Table 2 shows the inter-correlations of various trials of the same test, and Table 3 shows a factorial analysis of the 1st and 4th trials.

Improvements in Achievements

An inspection of Table 1 shows an increase in both means and S.D.'s with practice for all tests. In general, simpler intellectual tasks show smaller relative increases. Thus, aiming and tapping tests have a median increase in raw scores of 22% in four trials; that of feature and maze tests is 60%. Furthermore, hard maze and feature tests show greater relative increases than do easy forms. From subjective analyses above it seems clear that improvement on these trials depends upon selection and formation of routine procedures. When a routine pattern is well practiced, as in tapping with the dominant hand, little improvement is shown.

It also appears that the S.D.'s increase proportionately less than the means, so that the coefficients of variation ($C.V. = \frac{S.D.}{Mean}$) become smaller with practice. This table illustrates a general rule that tests showing smaller relative increase also have the smallest initial C.V.'s and the smallest changes in C.V.

These results lead one to examine the zeros of the scales. Obviously the raw score zeros do not represent just no ability, hence the true zero will be below the zeros of these scales. Thurstone (3) has postulated a true zero as that point where the dispersion of a normal group is zero. Our data here represent age ranges that are too small to justify an application of his formula, but on the basis of previous work, (4), it seems that all of these tests tend to have similar C.V.'s at approximately 14.7, when normal samples of various ages are

TABLE 2
Michigan Non-Verbal Series, Amounts Correct; Product-Moment Correlations
Between Trials of Equivalent Tests Uncorrected for Attenuation

Test	Trials	Trials			e^2	
		2	3	4	Begin	End
1	1	.531	.557	.457	.720*	
Aim Left	2		.676	.745		
1½ mm.	3			.793		.372
2	1	.495	.374	.290	.757	
Aim Left	2		.552	.451		
4 mm.	3			.782		.390
3	1	.397	.541	.443	.844	
Aim Right	2		.587	.633		
1½ mm.	3			.772		.406
4	1	.479	.268	.231	.762	
Aim Right	2			.569	.531	
4 mm.	3			.581		.663
5	1	.369	.303	.243	.864	
Tap Left	2		.374	.256		
10 sec.	3			.438		.808
6	1	.421	.414	.336	.822	
Tap Right	2		.346	.272		
10 sec.	3			.312		.902
7	1	.532	.583	.421	.716	
Easy	2		.591	.692		
Feature	3			.614		.624
8	1	.426	.374	.375	.818	
Medium	2		.473	.323		
Feature	3			.361		.870
9	1	.233	.211	.257	.945	
Hard	2		.352	.313		
Feature	3			.428		.816
10	1	.726	.592	.592	.472	
Easy	2		.690	.647		
Maze	3			.881		.220
11	1	.504	.513	.602	.747	
Medium	2		.494	.501		
Maze	3			.534		.716
12	1	.224	.280	.226	.950	
Hard	2		.351	.182		
Maze	3			.534		.716
Median	r_{12}	.450	r_{34}	.560	.796	.688

* Calculated from equivalent-form retests one day apart.

measured from true zero points calculated by Thurstone's method. The differences seen here in C.V.'s are therefore probably due largely to differences in raw score zero points, which, in turn, are related to the complexity of the task.

Changes in Self Correlation

Table 2 shows that during practice correlations between equivalent forms of a test tended to be larger between successive trials. The median r_{12} is .450, while the median r_{34} is .560 for all tests. The increases in correlation were greatest in aiming, hard feature, and hard maze tests, and smallest in tapping and medium feature and medium maze tests. The biggest correlations between 1st and 4th trials were found in aiming, medium and easy feature tests, and easy maze tests. There was a slight tendency for tests which had lower initial self correlations to have the larger increases with practices. This was particularly true of hard feature and maze tests and aiming with the dominant hand at larger circles. It was not true of tapping with either hand. Tapping was the test which allowed the greatest number of short variations in timing, because of the short time interval used, 10 seconds.

Factorial Analyses

The correlation matrices of the 1st and 4th trials were analyzed by Thurstone's (5) center-of-gravity method, using lengthened vectors. No effort was made to impose orthogonality among the planes, but every dimension was well represented.

The structure which appears from the five factors shown in Table 3 is one of the clearest seen from such data. Four planes, A, B, C, and D are very well defined, and E is fairly well delineated. All may be considered to be experimentally uncorrelated in the group tested.

An inspection of Table 3 shows that the total variances which were due to common factors, h^2 , were usually larger in the fourth than in the first trial. This means that group factors were more important after practice than before. We are not justified in claiming that the factors listed in both trials are the same. This claim could only be made if both trials had been included in a single factorial analysis. However, it seems probable from the subjective reports earlier in the article that the important changes in processes which accompanied practice on various tests can be described in terms of the factor loading changes found in Table 3. The two factorial analyses will be compared, and the factors tentatively named.

TABLE 3
Effects of Practice on Factorial Analyses (Greene, 1938,
Michigan Non-Verbal Series)

Test	<i>Form A. 1st Trial</i>						<i>Form D. 4th Trial</i>					
	A	B	C	D	E	h^2	A	B	C	D	E	h^2
1. Aim Left, Small	.10	.73	.00	.07	.00	.55	-.05	.46	-.08	.79	-.01	.83
2. Aim Right, Small	.02	-.09	.35	.69	.01	.61	.25	-.34	.01	.51	.06	.44
3. Aim Left, Large	.08	.72	.54	.12	.00	.83	.04	.34	.07	.75	.03	.68
4. Aim Right, Large	.08	-.03	.30	.64	.06	.51	.04	-.04	.41	.58	-.03	.51
5. Tap, Left	.01	.14	.57	.03	.01	.35	.05	.66	.40	-.01	.07	.60
6. Tap, Right	-.06	-.05	.56	.09	.02	.33	.01	.03	.76	-.03	-.07	.58
7. Feature, Easy	.47	.12	.34	.14	.12	.38	.73	-.11	.18	.04	.23	.63
8. Feature, Medium	.46	.08	-.11	.12	.19	.38	.70	.01	-.01	-.02	.05	.49
9. Feature, Hard	.69	-.06	.16	.05	.02	.50	.72	.12	-.09	-.06	-.16	.55
10. Maze, Easy	.04	.29	.06	.00	.68	.55	.07	.37	.05	-.01	.75	.58
11. Maze, Medium	.02	.32	.37	-.10	.71	.76	.00	.07	.04	.02	.71	.66
12. Maze, Hard	.09	-.19	.11	-.02	.39	.21	-.03	-.08	-.12	-.05	.75	.58

- A. Feature Discrimination (isolation disconnected visual patterns)
 B. Ambidexterity or persistence
 C. Tapping and easy aim. Both hands. (Speed of ballistic movement)
 D. Aim. Both hands. (Precision of movement or rhythm)
 E. Maze solution. (Trail finding)

On the first trial the left hands have a very high loading in factor B, and small loadings in D. In the 4th trial these loadings are nearly reversed. The right-hand performances in both trials of large and small circles are nearly zero in B and very high in D. Thus the left hands are seen to change toward the right-hand patterns. This change was also noted in the subjective analysis. Two of the aiming tests, Right small, and Left large, show a reduction with practice in factor C, which is closely identified with speed of tapping. Aiming, Left small, shows none of factor C in either trial, while aiming, Right large, shows moderate amounts of C in both trials. This finding corresponds to observation that ballistic movements were usually reduced with practice except in the case of aiming, Right large. There precision had been developed to a point where nearly automatic movements were effective.

Tapping tests with the left hand show small amounts of factor B and large amounts of C on the first trial. Tapping Right shows no B but a large C loading. On the 4th trial Tapping Left has a larger amount of B factor than formerly and a smaller amount of C, while Tapping Right has no B and more C than formerly. The tapping factor patterns of the two hands, therefore, are shown to be less alike

after practice than before, although tapping performance was more alike as practice continued. This analysis indicates that after practice the variations in tapping of the two hands were apparently related to factor B more than the variations in aiming.

In the 1st trial all the feature tests are weighted most heavily by factor A, but they also show small amounts of other factors, particularly C. In the 4th trial the Feature Discrimination tests are all heavily loaded with factor A and show no other significant loadings. A very small amount of factor C, which corresponds to speed of rough marking, is still found in the easiest test. With practice the easiest feature discrimination test becomes in its factorial patterns more like the hardest test. From the subjective analysis it appeared that practice eliminated unnecessary movements by establishing regular methods of comparison.

The maze tests give much the same factorial picture. In the 1st trial they are all heavily loaded with E but they also have small amounts of B and C. In the 4th trial they are all heavily loaded with factor E and show no other significant loadings.

The identification of the statistical factors with behavior patterns is justified to the degree that competent judges agree upon the analysis of processes. Since in this case the tests were fairly well controlled and limited to particular skills, the subjective identifications are fairly clear, with one exception, factor B. Factor B may be allied to ambidexterity, since it appears principally in left-hand Aiming and Tapping, and easy Maze tests, performed by persons who were right-handed in 90% of the cases. Factor B may also be an indication of persistence in a speed trial. Factor A is a form of non-verbal induction using visual patterns. It is found only in the Feature Discrimination Tests. Factor C corresponds most clearly to speed in a series of small ballistic movements, as in tapping. Factor D is identified with speed of precision of small hand and eye movements, coordinated as in aiming. Factor E is found only in the pencil mazes. It is therefore identified with isolation of a visual path in a connected visual pattern.

Discussion

Evidence of Systematic Variations

Systematic changes in behavior may be described as fairly persistent changes in patterns of response to standard stimuli. Such changes are often attributed to different combinations of various mechanisms, but they may be changes in energy only. Random changes are usually described as changes which persist only for very

short periods, and hence do not affect the person's usual mode of response. Systematic variations would be indicated by changes in group norms, while random changes might be present if there were no changes in group norms.

The subjective reports above indicated that some of the tests showed more evidence of systematic changes than others. The two tests of aiming with the right hand and tapping with each hand were reported to have undergone only small changes in procedure with practice, while in all the other tests large changes occurred. The short tapping and aiming tests were reported to fluctuate more than the rest, because of difficulties in starting and stopping promptly and because of variations in energy or willingness to work.

In the statistical data a systematic change is indicated when, in several trials of a test, individuals' scores show a series of changes in a given direction. Hence systematic changes would be indicated by a change in the mean scores in Table 1, a change in correlations between trials in Table 2, or a change in factor loadings in Table 3.

Table 1 shows that all tests changed in mean scores, but only small variations were found in Aiming, Right small. Each of the other tests showed at least 20 per cent improvement over the first score.

Table 2 shows that only five tests showed small changes in the size of correlation coefficients between adjacent trials. These tests include Aiming, Right large, Tapping, Left and Right, Medium Feature Discrimination, and Medium Pencil Maze. Hence systematic changes were probably present in the other test situations.

Table 3 shows only one test, Aiming, Right large, which has practically the same factorial loadings in the 1st and 4th trials. One must conclude that only this test shows no marked evidence of systematic changes. There may have been systematic variations present in performance, but none appear in this analysis of scores. All the other tests have rather marked changes which probably indicate systematic variations in individual patterns of response.

The Most Probable Limits of Random Variation

Thurstone (5), Garrett (6), and Guilford, (7) have used e^2 in the formula

$$e^2 = 1 - r^2_{11}$$

to indicate the proportion of the total variance which is called random, in a special situation where it can be assumed that no systematic changes have taken place in two trials 1 and I of a test. Since this assumption has been shown to be untenable in most of the tests under

consideration, what value may be assigned to random variance? No direct answer appears to this question. If one accepts the hypothesis that random factors will be at least as great at any point during the practice as they will be at the end of a long practice series, then the self correlation at the end of practice will indicate a minimum e^2 . This is a reasonable assumption in almost any series, because individual consistency usually improved with practice. For the same reason the maximum true e^2 for any point in a practice series will probably not be greater than that actually found somewhere near the beginning of the series. The maximum e^2 will usually be less than that indicated, for systematic changes, which eventually may become fairly permanent, can never be discerned from two adjacent trials alone. The e^2 's shown in Table 2 are therefore probably higher than they should be, with the exception of Test 4, which seems to have reached a fairly stable pattern for the persons tested.

The Significance of the Quantitative Data

The quantitative data present several significant findings since they represent behavior patterns of a fairly large group of adolescent boys. First, the relative mean and standard deviation increases are large enough to make it necessary to compensate for practice effects in practical comparisons of individuals at various stages of practice. Second, the stability of factor loadings of one aiming test and the lack of stability on the other three aiming tests and on the tapping tests are of interest to those who seek measures of lateral dominance. Third, the factorial analyses are of interest to those who are approaching the understanding of mental organization from this angle. The results show quantitative indicators of changes in mental organization which take place during practice, and also indicate that most of the changes which were reported subjectively can be well represented as the resultants of five factors. The clarity with which these factors are isolated statistically gives promise that, from tests which are somewhat better controlled than these, very pure measures of abilities may some time be made which will be relatively free from factorial changes during practice. The nature of the five factors is also of interest. Evidence is presented here that ballistic and precision movements are unrelated and that isolation of a trail from a maze and isolation of a feature in one of several faces are two unrelated forms of selective activity.

Summary

Six motor tests and six non-verbal tests of observation and comparison were applied four times in one week to 394 adolescent boys

enrolled in a trade school. Observations and subjective reports of changes with practice were compared with quantitative results. The more complex tests showed more improvement with practice, more reduction of C.V.'s, and less predictability than did the less complex tests. Subjective reports indicated more varieties of approach to the complex than to the easy tasks and more elimination of useless acts among the complex tests by the methodical habit forming.

The first and last trials showed unrelated factors which are tentatively identified with:

- a) Isolation of unique visual pattern from series of disconnected patterns. (Feature discrimination)
- b) Ambidexterity. (Left-hand aiming and tapping, and easy maze solution)
- c) Speed of ballistic movements. (Tapping)
- d) Speed of precision movements. (Aiming)
- e) Isolation of visual path from larger pattern. (Maze)

All except one test showed marked shifts in the size of factor loadings with practice. The shift was usually toward a larger loading on one factor and smaller loadings in all the rest. The presence of one test (Aiming, Right hand, Large circles) which maintained one factor pattern throughout the practice series, points to the possibility that more tests with constant factor patterns may be constructed.

The results lead to two suggestions. (a) Corrections for attenuation should be limited to situations where no systematic changes occur between trials as shown by factor loadings and subjective reports. (b) Since reliability coefficients may be indicative of both systematic and random changes, they should not be used routinely in corrections for attenuation. The probable limits of random variation can be described from a long practice series.