# THE RELATION BETWEEN ABILITIES AND IMPROVEMENT WITH PRACTICE

HERBERT WOODROW

University of Illinois

The chief purpose of the present analysis was to determine in the case of a number of tests the effect of practice on the factor-loadings. Changes in the factor-loadings of a test-performance would indicate a change in the degree to which the performance depended upon the various abilities possessed by the subjects; and, if such changes occur, it should be of value to know their nature. For example, it is of interest to know whether with practice performances show an increased dependence upon a speed factor, or, in case such a factor were found, upon "*g*."

It was also hoped to identify the factors upon which gain scores depend, and thus incidentally determine whether the gain scores in different performances depend upon the same factor or factors. In this respect the investigation was not very successful, but, nevertheless, some rather interesting conclusions concerning gain scores appear to be indicated. Gain scores are, of course, not independent scores, since they are differences between final and initial scores. As a result, unless both final and initial scores are in terms of the same units, the gain scores will be distorted. The units here used, with one exception, are raw score units. These units have significance, but there are other units which may be more significant. While, for example, the number of units of work done in a given time constitutes a valid, practical, and self-descriptive measure, an increase in the number of units done from ninety to ninety-five may represent a much greater increase in ability than an increase from forty to forty-five, or *vice versa*. Were the raw scores transformed into units corresponding to equal steps in ability, in other words, subjected to absolute scaling, there is little doubt but that the conclusions here reached concerning factors in gain scores would be considerably modified.

The data consist of the raw scores made by fifty-six subjects who completed thirty-nine days of practice in each of seven tests, all of which were given as group tests, and who, in addition, took a number of end tests, including intelligence tests, either before or after practice. In none of the practice tests, with the exception of the speed of making gates, was the same form used two days in succession. In the speed test, only one form is feasible. The scores made in all tests, including

in the case of the practice tests initial, final, and gain scores—a total of thirty-three variables—were intercorrelated and the correlational matrix subjected to a factor analysis by Thurstone's centroid method. The factors were then rotated by the graphic method so as to maximize the number of insignificant factor loadings.

The seven tests used as practice tests were as follows:

1. *Horizontal Adding.*—Thirty problems in addition, each problem consisting of adding six numbers, varying in length from three to seven places and arranged in a horizontal line. Ten forms. Time, ten minutes. Score, the total number of correct digits in the correct place in the answers.

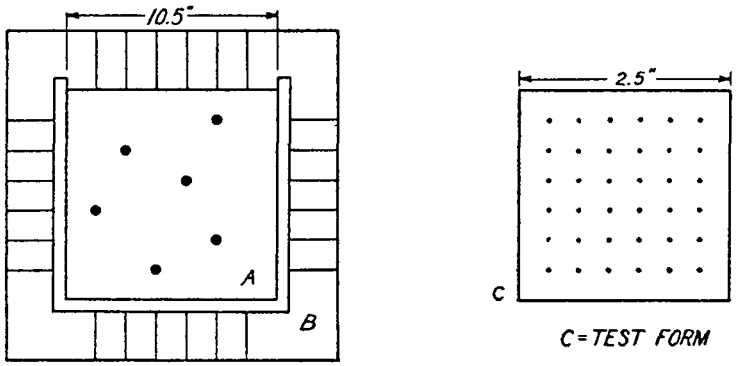


FIG. 1.—Modified spot-pattern test.

2. *Substitution.*—Writing a digit under each letter of a page of evenly spaced letters in accordance with a key list of paired digits and letters. The test sheets contained thirty lines of capital letters. Each line was made up of eight letters BFHKMTWZ, double spaced and arranged in irregular order. A new key was used each day, but otherwise the test sheet was the same throughout. Time, ten minutes. Score, number correct minus number of errors.

3. *Spot-pattern Test, Modified.*—On twenty-four exposure cards, ten and five-tenth inches square, were stamped black disks, one-half inch in diameter, in irregular patterns varying in number from four to nine, four of each number. The spots, or disks, were so placed that they always fell at one of the intersections of an imaginary cross-section sheet composed of squares one and five-tenth by one and five-tenth inches. The cross-section lines were not drawn on the stimulus-cards but were indicated by short lines drawn on a framework in which the cards were exposed (see Fig. 1). Each card was exposed to the subjects as a group for fifteen seconds. The subjects were

provided with response sheets on which were printed twenty-four rectangular forms. Each of these forms was one and five-eighth inches square and consisted of a set of thirty-six dots representing the intersections of imaginary cross-section lines. Each of the exposed patterns had to be reproduced by encircling the proper dots on one of the response-forms, *after removal* of the stimulus-card. The time allowed for reproducing each pattern was thirty seconds. Thirty-six forms were used, that is, thirty-six different sets of twenty-four stimulus-patterns each. The response-sheets were of course the same throughout. The original score on the spot-pattern test was the percentage of errors. The assumption was then made that goodness of performance varies (inversely) with the $\sigma$ value of these per cent scores, and such $\sigma$ values were the scores used.

4. *Anagrams.*—The task consisted in rearranging letters to make words. The test-sheets were composed of one hundred twenty sets of letters, arranged in four columns of thirty each. The number of letters per word in the columns increased from four in the first to seven in the last. The thirty-nine completely different forms required a list of five thousand seventy words. Time, ten minutes. Score, total number of words correct.

5. *Cancellation with Multiple Instruction.*—The test blanks provided forty-nine lines of all letters of the alphabet printed as capitals with uniform spacing in eleven point, Intertype Scotch, with approximately thirty-eight letters to a line. Nine forms were used. When required, several sheets of the same form were distributed for one day's practice. The instructions, modelled after those used by Philip,[1] were as follows:

"1. Draw a line through each vowel (A, E, I, O, U) which comes between two consonants, *i.e.*, which stands alone between two consonants.

"2. Where there are two vowels and nothing else between a pair of consonants, draw a line through the second one.

"3. Where there are three or more vowels between two consonants, do nothing.

"Work as rapidly and as accurately as you can. Your score depends upon speed and accuracy."

Time, ten minutes. Score, number of correct cancellations minus number wrong.

6. *Length Estimation.*[2]—The subject was required to estimate the length of the right-hand segment of a black, one-meter rod divided by a white pointer, as a percentage of the total length of the rod. The dividing pointer was given one hundred different settings at each practice period. Each list

---

[1] Philip: *The Measurement of Attention.* Catholic University of America. Studies in Psychology, Vol. I, No. 2, 1929.

[2] This test was devised and given by F. L. Ruch. He also provided the material used in the anagrams test and participated in the conduct of the experiment in various ways.

of one hundred settings was used only once. Score, total of arithmetical deviations of estimated percentages from the correct percentages.

7. *Making Gates.*—A speed test in which the subjects were to make "gates" consisting of four horizontal and one diagonal line in each square of a page on which was printed a rectangle seven by nine inches divided into one thousand eight squares. Time, ten minutes, divided into five two-minute periods separated by rest periods of one minute. Score, number of gates (five lines each) completed.

Since a considerable number of tests is required to obtain a meaningful set of reference abilities, other tests, termed end-tests, were given, both before and after the thirty-nine days devoted to practice. Inasmuch as the factors in improvement were unknown at the time of testing, these additional tests had to be chosen rather blindly. Had the outcome been known in advance, a longer and probably better list of tests would have been used.[1]

The end-tests were as follows:

1. *Thorndike CAVD, Levels M, N, O, P, and Q, Form* 2.—Time allowed, two hours, forty minutes. Most of the subjects claimed they had answered all the questions they were able to answer considerably before the booklets were collected, and were allowed to cease working on them. This test was given after the practice sittings were ended.

2. *Otis Group Intelligence Scale, Advanced Examination, Forms A* and *B.*—Only six of the tests were given; namely, those labelled Directions, Proverbs, Arithmetic, Geometric Figures, Similarities, and Narrative Completion. The times allowed were not identical with those specified in the manual of directions, and were shorter with form *B*, given second, then with form *A*, given first. The twelve scores obtained with the two forms were summed into one total score. Both forms were given before the practice sessions began—form *B*, by three days, and form *A*, by ten days.

3. *Analogies.*—Two forms, especially prepared, each containing sixty items, with the correct answer to be indicated by underlining one of five given choices. One form was given before, and one form after practice. Time, five minutes, thirty seconds. Score, number right.

4. *Form Analogies, and* 5, *Artificial Language.*—Tests 4 and 5 were the ones contained in the Psychological Examination published by the American Council of Education. Two scores were obtained with each test. The first score was the average made on two forms, 1930 and 1931, given ten days apart, before practice, and the second score was that made on the 1932 form, given after practice.

[1] A second experiment has now been completed with a larger number of subjects, a greater number of practice periods, a larger and, it is believed, better set of "end" tests. In the factor-analysis of the data, gain scores will be omitted.

6. *Thurstone's Categories Test.*—Marking each of six words as 1 or 2, to indicate in which of two categories they belong, the nature of the two categories being indicated by two adjacent lists of four words each. A total of thirty-two sets of six words each, to be so marked. Fore-practice was provided. The test was given before practice, in two segments, separated by ten days. Total time, exclusive of fore-practice, fourteen minutes. Score, number of words correctly marked.

7. *Mental Multiplication.*—Twenty-five problems. Answers to be written upon signal "write." Time of ten seconds allowed for the five two-by-one place multiplications and fifteen seconds for the twenty three-by-one place multiplications. Score, number of digits correct. The score used is the average of two scores from two forms both given before practice.

8. *Speed of Making Crosses.*—The test sheet was the same as that for the "making gates" practice test. One cross to be placed in each square. Time, six minutes, preceded by one minute of fore-practice followed by one minute of rest. Score used, average of two trials, both given before practice.

9. *Three-digit Cancellation.*—The subjects were instructed to cancel every 2, 4, and 9 on a printed page consisting of the digits 0 to 9, inclusive, arranged in irregular order and evenly spaced. Time five minutes.

It is extremely important in studies of practice to have highly reliable initial and final scores. As is well known, unless the scores are highly reliable, the correlation between initial scores and gains appears less positive or more negative than the true correlation. High reliability of initial and final scores was secured, in part, by using relatively long tests (no test requiring less than ten minutes), and also by amalgamating the scores made at several sittings. What is here termed the initial score is the average of the first several days of practice, preceded by several minutes of fore-practice and by a number of mental tests given before the experiment proper began in order to accustom the subjects to group-test procedure. What is called the final score is also the average of several scores. As a result, the Spearman-Brown reliability coefficients are all over +.90 and average +.94, and are as high for initial scores as for final scores. The reliability of the various scores, as well as the mean and the standard deviation of their distribution, is shown in Table I.

Twenty-one of the total of thirty-three variables which were intercorrelated consisted of the three scores, initial, final, and gain scores, from each of the seven practice tests; the other twelve were scores from the end tests. There were only nine different end tests, but in the case of three of these tests; namely, artificial language, form analogies, and verbal analogies, scores obtained from two different forms, given before and after the practice sittings, were retained as

separate variables. The five hundred twenty-eight coefficients of correlation between the pairs of these thirty-three variables were

TABLE I.—DATA CONCERNING PRACTICE TESTS
$r$ = Spearman-Brown reliability coefficient

| Test | Initial score | | | | Final score | | | |
|---|---|---|---|---|---|---|---|---|
| | Days pooled | $r$ | Mean | $\sigma$ | Days pooled | $r$ | Mean | $\sigma$ |
| Horizontal adding | 1st 5 | .938 | 47.1 | 11.3 | last 3 | .976 | 85.6 | 21.2 |
| Substitution | 1st 5 | .944 | 395 8 | 61 3 | last 3 | .942 | 487.2 | 96.1 |
| Spot-pattern | 1st 5 | .949 | +.458 | .397 | last 3 | .926 | +1 763 | .603 |
| Length | 1st 5 | .906 | 103.0 | 32 2 | last 3 | .920 | 88.0 | 31.4 |
| Cancellation | 1st 2 | .954 | 280.2 | 59 9 | last 2 | .960 | 576.0 | 112 1 |
| Making gates | 1st 3 | .965 | 506.0 | 79 7 | last 3 | .938 | 617.5 | 79 0 |
| Anagrams | 1st 5 | .938 | 34 1 | 7 4 | last 3 | .943 | 51.2 | 8.2 |

calculated by the Pearson product-moment method. As a sufficiently accurate estimate of the communalities required for the diagonal entries in the correlational matrix, the highest correlation of each variable with any of the others was used. A new estimate made in the same way was used in each matrix of residual correlations. The nine factors obtained by the centroid method[1] are shown in Table II.

The test numbers in Table II (as also in Table IV) stand for the following test scores:

1. Horizontal Adding, initial
2. Horizontal Adding, final
3. Horizontal Adding, gain
4. Substitution, initial
5. Substitution, final
6. Substitution, gain
7. Spots, initial
8. Spots, final
9. Spots, gain
10. Multiple Cancellation, initial
11. Multiple Cancellation, final
12. Multiple Cancellation, gain
13. Relative Per Cent Length, initial
14. Relative Per Cent Length, final
15. Relative Per Cent Length, gain
16. Speed, gates, initial
17. Speed, gates, final
18. Speed, gates, gain
19. Anagrams, initial
20. Anagrams, final
21. Anagrams, gain
22. Artificial Language (A, before practice)
23. Artificial Language (B, after practice)
24. Form Analogies, A.
25. Form Analogies, B.
26. Verbal Analogies, A.
27. Verbal Analogies, B.
28. Thorndike CAVD
29. Average 6 Otis Form A. and 6 Otis Form B.
30. Categories
31. Cancellation, 3-digit
32. Arithmetical Problems
33. Speed, making crosses

[1] The procedure followed was that described by Thurstone in *The Vectors of Mind*, 1935, Ch. III.

TABLE II.—ORIGINAL FACTOR LOADINGS

The factors are designated by the roman numerals I to IX.

$h^2$ = proportion of the total variance of the test due to the nine common factors.

| Test | I | II | III | IV | V | VI | VII | VIII | IX | $h^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | .588 | −.180 | −.145 | .202 | −.092 | .125 | .047 | − 256 | −.296 | 619 |
| 2 | .563 | −.364 | −.288 | 386 | −.282 | .268 | −.256 | −.066 | −.100 | 913 |
| 3 | .347 | −.343 | −.262 | .314 | −.317 | .192 | −.361 | .098 | .103 | .693 |
| 4 | .482 | −.337 | −.215 | −.220 | .336 | .083 | 086 | .183 | −.160 | .627 |
| 5 | 466 | −.632 | 022 | −.358 | −.134 | .268 | .226 | −.187 | − 151 | .944 |
| 6 | .190 | −.449 | .189 | −.230 | −.381 | .224 | .110 | −.262 | −.015 | .603 |
| 7 | .656 | .242 | −.249 | − 158 | .151 | .064 | 069 | −.136 | .112 | .639 |
| 8 | .529 | .288 | −.507 | −.361 | −.127 | .313 | .205 | −.072 | .241 | .970 |
| 9 | .130 | .144 | −.421 | − 289 | −.322 | .291 | .117 | .074 | .132 | .523 |
| 10 | .489 | −.263 | − 172 | − 099 | .292 | −.344 | .284 | 117 | −.167 | .674 |
| 11 | .614 | −.500 | 296 | −.107 | −.081 | −.312 | 209 | .176 | .196 | .948 |
| 12 | .499 | − 308 | .580 | .129 | −.246 | 071 | .265 | .322 | .230 | 989 |
| 13 | .355 | .386 | −.449 | − 245 | 018 | − 115 | − 084 | .444 | −.209 | .798 |
| 14 | 319 | .130 | − 160 | − 495 | −.174 | −.516 | − 370 | .123 | .179 | .870 |
| 15 | − 091 | −.184 | .177 | − 238 | −.066 | −.323 | −.238 | − 347 | .268 | .488 |
| 16 | .506 | −.241 | −.151 | .331 | .456 | −.116 | −.123 | −.117 | .231 | .750 |
| 17 | .649 | −.304 | − 044 | −.256 | .233 | .068 | − 249 | .126 | −.196 | .756 |
| 18 | .045 | −.009 | .003 | −.477 | −.208 | .439 | − 164 | .136 | −.441 | 706 |
| 19 | .414 | .274 | −.275 | .435 | −.109 | −.445 | 251 | −.264 | −.250 | 916 |
| 20 | 429 | .147 | −.343 | .547 | −.078 | −.296 | 067 | − 143 | −.035 | .742 |
| 21 | .069 | −.107 | −.139 | .181 | −.011 | .094 | −.154 | .043 | 234 | .158 |
| 22 | .609 | .302 | .286 | .144 | .209 | −.027 | −.077 | −.033 | −.074 | .622 |
| 23 | .687 | .306 | 197 | .123 | .096 | .128 | .225 | .059 | .063 | .703 |
| 24 | .681 | .276 | .298 | −.071 | −.077 | .019 | −.125 | −.140 | −.116 | .689 |
| 25 | .672 | .234 | .097 | −.207 | .109 | .029 | −.207 | .060 | .120 | .632 |
| 26 | 576 | .440 | 292 | 317 | .227 | −.015 | − 122 | −.160 | .081 | .810 |
| 27 | .517 | .446 | .315 | .198 | .336 | −.019 | −.116 | −.123 | −.041 | .748 |
| 28 | .648 | .335 | .300 | .123 | −.282 | −.184 | −.091 | −.071 | − 108 | .776 |
| 29 | .738 | .433 | .395 | 060 | .023 | .030 | .010 | − 060 | −.025 | .900 |
| 30 | .528 | .336 | 345 | .279 | .407 | .129 | .063 | .113 | .194 | .825 |
| 31 | .469 | −.425 | .232 | −.043 | 045 | −.064 | .225 | .344 | − 315 | .729 |
| 32 | .399 | .333 | −.144 | .094 | −.289 | .138 | .195 | .199 | .021 | .480 |
| 33 | .450 | − 387 | − 165 | − 045 | 378 | −.186 | 018 | − 113 | 152 | 595 |

RESIDUALS AFTER REMOVAL OF NINE FACTORS

| Magnitude | Frequency | Magnitude | Frequency |
|---|---|---|---|
| +.270 to + 251 | 1 | − 010 to −.029 | 85 |
| +.250 to + 231 | 0 | −.030 to − 049 | 53 |
| + 230 to + 211 | 1 | − 050 to −.069 | 38 |
| +.210 to + 191 | 1 | −.070 to − 089 | 25 |
| +.190 to + 171 | 4 | − 090 to −.109 | 10 |
| + 170 to +.151 | 0 | −.110 to − 129 | 9 |
| +.150 to +.131 | 3 | − 130 to − 149 | 4 |
| + 130 to + 111 | 8 | − 150 to −.169 | 5 |
| +.110 to + 091 | 19 | −.170 to − 189 | 1 |
| +.090 to +.071 | 18 | − 190 to −.209 | 5 |
| +.070 to +.051 | 32 | | |
| +.050 to +.031 | 49 | −.250 to −.269 | 2 |
| +.030 to +.011 | 70 | | |
| +.010 to −.009 | 84 | −.350 to −.369 | 1 |

Mean Residual = − 004; $\sigma_{dis}$ = .066

The transformation matrix was calculated by the graphic method. All pairs of axes were rotated until, so far as could be determined by inspection, no further rotation of any of the possible pairs would produce an increase in the number of insignificant factor loadings. Sixty-two rotations of pairs of axes were made. Since there were only fifty-six cases, it is clear that the $\sigma$ of any loading must be large. Since the $\sigma$ of an original correlation of zero when $n$ is fifty-six is .13, in rotating axes all loadings of less than twice this amount, *i.e.*, $\pm .26$, were regarded as small and probably insignificant. Only one negative loading in excess of this magnitude remained after the rotation of axes; namely, the loading of $-.287$ on the part of variable eighteen, gain in speed, with factor VI. Some negative loadings are necessitated by the fact that negative correlations occur in the original matrix.

TABLE III.—TRANSFORMATION MATRIX

| | I | II | III | IV | V | VI | VII | VIII | IX |
|---|---|---|---|---|---|---|---|---|---|
| *a* | .533 | .319 | .294 | .227 | .416 | .383 | .198 | 206 | .282 |
| *b* | .541 | −.406 | .261 | −.316 | −.491 | .219 | −.065 | .170 | −.234 |
| *c* | .480 | .474 | −.472 | −.346 | −.143 | −.302 | .165 | −.252 | .011 |
| *d* | .201 | −.001 | −.360 | .567 | −.147 | .376 | −.420 | −.372 | −.182 |
| *e* | .285 | −.518 | −.174 | −.138 | .693 | −.220 | −.118 | −.148 | − 194 |
| *f* | .187 | −.022 | .472 | .154 | −.121 | −.493 | −.394 | −.340 | .432 |
| *g* | −.166 | .376 | .310 | −.494 | .219 | .311 | −.422 | −.299 | −.280 |
| *h* | .095 | .250 | −.112 | .086 | .048 | −.297 | −.547 | 693 | −.197 |
| *i* | .068 | .181 | .366 | .347 | .000 | −.315 | .322 | −.132 | −.697 |

The transformation matrix calculated from the rotations of pairs of axes is given as Table III. Each column of the matrix may be regarded as an equation. For example, if the original nine factor loadings (Table II) of any test be designated $a, b, \cdots i$, then the new loading with Factor I of that test will equal $.533a + .541b \cdots + .068i$. Similarly, the second column of the transformation matrix gives the equation by which the loadings of any test with the rotated Factor II may be obtained from the original loadings with all nine factors. In this way, that is, by multiplying the original factor matrix (Table II) by the transformation matrix (Table III), the transformed factorial matrix (Table IV) is obtained (in Thurstone's notation, $FG = V$). The transformation is orthogonal, and, therefore, the transformed factors (or rotated axes) remain orthogonal. Only the results obtained after this transformation, or rotation, should be considered as representing the outcome of the analysis. The data given in Tables II and

III merely represent necessary steps, and are presented solely because they permit further computations, or further rotations of axes, should one care to make them.

TABLE IV.[1]—FACTORIAL MATRIX AFTER ROTATION

| Test | I | II | III | IV | V | VI | VII | VIII | IX |
|---|---|---|---|---|---|---|---|---|---|
| 1 | .132 | .137 | .132 | 239 | .243 | .448 | .006 | −.129 | .484 |
| 2 | .042 | .200 | .134 | .727 | .111 | .270 | −.036 | −.022 | .497 |
| 3 | −.042 | .194 | .084 | .752 | −.014 | .057 | 008 | .106 | .272 |
| 4 | .032 | .062 | .163 | .049 | .680 | .001 | −.088 | .222 | .276 |
| 5 | − 209 | .490 | .276 | −.026 | .469 | −.031 | .155 | −.133 | .565 |
| 6 | −.208 | .497 | .163 | −.005 | .026 | −.077 | .244 | −.232 | .405 |
| 7 | .368 | −.074 | .512 | .051 | .318 | .281 | .177 | .124 | .083 |
| 8 | .120 | −.026 | .930 | .042 | .121 | .166 | .077 | .190 | .091 |
| 9 | −.154 | .031 | .648 | .049 | −.143 | .000 | −.054 | .213 | .099 |
| 10 | −.012 | .144 | −.007 | −.083 | .684 | .356 | −.008 | .228 | .023 |
| 11 | .090 | .746 | −.067 | .077 | .511 | .119 | .261 | .158 | −.037 |
| 12 | .349 | .907 | −.046 | .104 | .151 | −.069 | −.052 | −.017 | −.022 |
| 13 | .158 | −.222 | .296 | −.039 | .087 | .226 | −.158 | .739 | .080 |
| 14 | .003 | −.002 | .105 | .006 | .031 | .096 | .608 | .685 | −.086 |
| 15 | −.166 | .044 | −.151 | −.019 | −.014 | −.098 | .634 | −.090 | −.117 |
| 16 | .268 | −.079 | −.036 | .481 | .598 | .192 | .134 | −.124 | −.105 |
| 17 | .229 | .090 | .053 | .157 | .567 | −.039 | .169 | .318 | .467 |
| 18 | −.043 | .009 | .198 | −.221 | −.133 | −.287 | −.085 | .266 | .658 |
| 19 | .126 | −.060 | −.008 | .061 | .032 | 945 | −.019 | − 019 | −.044 |
| 20 | .150 | −.057 | .027 | .388 | 065 | .744 | −.047 | −.005 | −.097 |
| 21 | .009 | −.002 | .072 | .378 | 024 | − 066 | .002 | −.021 | −.076 |
| 22 | .714 | .050 | −.025 | .002 | .172 | .243 | .100 | .038 | .106 |
| 23 | .676 | .246 | .274 | −.016 | .194 | .265 | −.071 | .006 | .019 |
| 24 | .623 | 183 | .114 | −.051 | .025 | .252 | .291 | .098 | .310 |
| 25 | .574 | .067 | .256 | .066 | .211 | .043 | .288 | .293 | .128 |
| 26 | .822 | −.046 | −.004 | .112 | .059 | .289 | .144 | −.113 | −.036 |
| 27 | .804 | −.123 | −.056 | − 038 | .126 | .224 | .112 | −.059 | .021 |
| 28 | .582 | .291 | − 006 | .020 | −.151 | .456 | .251 | .165 | .176 |
| 29 | .832 | .217 | .133 | − 089 | 040 | .291 | .164 | .049 | .135 |
| 30 | .839 | .069 | .048 | .063 | .250 | .048 | −.103 | −.095 | −.175 |
| 31 | .098 | .525 | −.190 | − 070 | .482 | .090 | −.187 | .211 | .294 |
| 32 | .275 | .198 | .399 | .078 | −.155 | .295 | −.210 | .213 | .023 |
| 33 | .012 | .038 | .046 | .208 | .692 | .121 | .237 | −.010 | −.018 |

[1] See explanatory note preceding Table II.

As a matter of fact, numerous further calculations besides those here described have been made. For example, it was found possible to obtain factors closely resembling the first five factors of Table V

by a procedure which utilizes quite different sign-changing rules than those of Thurstone's centroid method.[1]

From the transformed factorial matrix (Table IV), it may be seen that every test, with the exception of number 21 (improvement in anagrams), has a significant loading (.39 or over) with at least one factor and an insignificant loading (under .26) with at least as many as five factors. Only one variable, final substitution score, shows loadings over +.39 with as many as three factors.

To attempt to name the factors is hazardous since conventional names ordinarily apply to a total or complete operation, whereas a factor, unless it shows one or more loadings approaching unity, is only one abstract causal condition acting along with others in the determination of goodness of score in any whole operation. Moreover, since some of the variables such as the Thorndike and Otis scores represent composite scores, and since the number of tests was relatively small, there is little likelihood that the factors obtained represent truly "primary" abilities. Nevertheless, it appears desirable to attempt roughly to identify the factors. Consequently, it will be pointed out with what tests each of the factors shows the highest loadings, and thus in what sort of performances each factor is important.

*Factor* I.—Important in tests of intelligence or tests such as opposites and analogies that have been alleged to be good tests of "*g*."[2] That it is not "verbality" is indicated by the very low correlation with anagrams; yet the only non-verbal test which correlates as high as +.39 with this factor is that of form-analogies. An attempt was made to discover some method by which the loading shown by form-analogies with this factor could be reduced to insignificance. None was discovered.

*Factor* II.—Important in tests which have not infrequently been designated tests of attention. The highest correlations of this factor are with final and gain score (11 and 12) of the Philip's multiple instruction letter cancellation test, one of Philip's battery for measuring attention. Other tests with high loadings are 3-digit cancellation (31) and substitution final and gain score (5 and 6). On the other hand, the correlation with horizontal adding, which was also one of the tests of attention devised by Philip, is insignificant. This last test appears to be more of a computation or numerical test, than an atten-

[1] Woodrow, H. and Wilson, L. A.: "A simple method of approximate factor analysis." *Psychometrika*, Vol. II, 1937, pp. 245–258.

[2] Spearman, C.: *The Abilities of Man*, 1927.

tion test, in the case of the present group of subjects. Possibly this factor could be termed speed of perception of detail, or, perhaps better, conceived as a factor in tests of analytic reaction (reactions to items which need to be attended to separately).

*Factor* III.—The only certainly significant correlations are with the spot-pattern scores. Final spot-pattern score shows the high factor loading +.930.

*Factor* IV.—Possibly a numerical factor, since the highest correlations are with final and gain scores in horizontal addition. The correlations of +.388 and +.378 with final and gain scores in anagrams, though low and possibly insignificant, make an interpretation of this factor somewhat hazardous. The correlation of +.481 with initial score in speed of making gates is also hard to explain.

*Factor* V.—Rather clearly a speed factor. The significant correlations are with speed of making gates, initial and final scores, speed of making crosses, Philip's cancellation, initial and final, digit-cancellation, and substitution, initial and final.

*Factor* VI.—Correlates +.945 with initial anagrams score, but also shows a correlation of +.448 with initial horizontal adding score. A puzzling relation between anagrams and horizontal adding thus appears in the case of two different factors, Factor IV being involved in the final and gain scores of both tests and Factor VI being particularly prominent in the initial scores of both tests.

Factors VII and VIII both pertain primarily to the test of estimation of relative length, Factor VII owing its existence largely to the correlation of the final and gain scores and Factor VIII to the correlation of the initial and final scores. This result illustrates the complication resulting from using three different scores, initial, final, and gain, derived from one practice test, in the same matrix of correlations.

Factor IX shows the highest loading in the case of gain in speed. Its interpretation may be connected with the meaning of the gain in speed scores, but it is not apparent that the other loadings throw any light upon that interpretation.

Certainly great caution should be exercised in drawing any final conclusions from the preceding analysis. On account of the small number of cases, and the approximations used as communalities, and the further fact that a merely graphic method of rotating axes has been used, the results should, no doubt, be considered as only a preliminary approximation to the truth. Nevertheless, certain facts stand out so decisively as to leave little doubt of the validity of certain general

conclusions which are of value in connection with a number of problems concerning practice.

1. Perhaps the most important fact established is that marked changes in factor loadings occur with practice. In the case of two of the tests, however, speed and anagrams, the changes are not enormous, and in the case of anagrams are possibly not significant, though the change from .945 to .744 in the loading with Factor VI and that from .061 to .388 with Factor IV are rather pronounced changes. While, then, it may not be established beyond doubt that practice *always* produces significant changes in factor loadings, there can be little doubt that such changes usually occur. Horizontal adding shows a change from +.239 to +.727 in the loading with Factor IV. Substitution shows a change in the loading with Factor II from +.062 in the initial performances to +.490 in the final performances. At the same time it shows a drop in the loading with Factor V from +.680 to +.469. The spot-pattern test shows a rise from +.512 to +.930 with Factor III. Philip's cancellation test shows an increase from initial to final score in the Factor II loading from +.144 to +.746. Estimation of relative length changes in loading with Factor VII from −.158 to +.608. Speed of making gates changes from +.481 to +.157 in its loading with Factor IV and from −.105 to +.467 in its loading with Factor IX.

That these large changes in the factor loadings are due to practice is indicated by the absence of such changes in the case of the three non-practice tests—artificial language, form-analogies, and verbal analogies—given twice, before and after practice. The two scores of each of these tests show variations in their factor loadings of only that magnitude which has here been considered unreliable, and the largest change shown in the loadings of any of these three tests with any of the factors; namely, the change from −.025 to +.274 in the loading of artificial language with Factor III, is far less significant than the changes shown by the tests in which practice was given.

It is interesting to observe that a recent analysis made by the Hotelling method[1] harmonizes well with the finding of marked changes in factor loadings, since it shows that decided changes in the weights of the various principal components resulted from a brief period of instruction concerning various helpful devices, this instruction being interpolated between the initial and final administration of the tests.

[1] Anastasi, A.: "The influence of specific experience upon mental organization." *Genetic Psychol. Monog.*, Vol. XVIII, No. 4, 1936, pp. 245–355.

In general these changes with practice in the factor loadings mean that the quantitative pattern of abilities determining goodness of performance changes with practice, *i.e.*, a performance after practice is likely to depend for its success more on one ability or less on another than it did initially. Such a change must mean a change in the mode of operation whereby the subject carries out the task he has been instructed to accomplish.

In a sense the task which the subject practices remains the same. There is no change either in the instructions, that is, in the task the subject is asked to perform, nor in the manner in which the experimenter scores the records made on the test-papers. If the goodness of the scores be regarded as determined by a *set* of coöperating but independently variable abilities, then practice may be regarded as a change in the conditions under which the various constituents of this set of abilities operate. A fixed amount of practice may be regarded as a fixed change in conditions; but this fixed change in the total constellation of conditions does not result in an equal increase in favorableness for the operation of all the coöperating determining factors.

2. There is no general tendency for the loading with Factor I or for the average $r$ with four "intelligence" tests to be larger in the case of final scores than initial scores. In fact, the tendency is rather in the opposite direction, though the changes are small. Only in the case of the Philip's cancellation test does final score have a higher positive factor loading than initial score. In the case of Factor V, also, a factor which is rather clearly a speed factor, the loadings tend to decrease with practice. In none of the seven practice tests does final score show a significantly higher loading with the speed factor than does initial score. In fact, the final score loading is smaller than the initial score loading for all tests except anagrams, in which case both the initial and final loadings are insignificant (+.032 and +.065). If, then, initial and final measures are equally reliable, as is here the case, it is not true that there is any *general* tendency for test-performances to become with practice more dependent upon or better measures of some supposedly common factor, such as "g," intelligence, or speed. One reason why test scores have sometimes been supposed to do so is probably the fact that as a rule previous investigators have not used initial scores of as high reliability as the final scores.

3. There is no sign of any general improvement factor, that is, a factor common to the gain scores of all the practice tests. It is

particularly noteworthy that Factor I, which somewhat resembles Spearman's "*g*" factor, shows little if any correlation with any of the gain scores. Only in one case, that of multiple-instruction cancellation, could the loading with this factor (+.349) possibly be regarded as significant. Factor I loadings resemble the average correlations of each variable with four variables which, taken together, may be regarded as measuring "intelligence" as it is commonly conceived. These four variables are the Thorndike *CAVD*, the pool of six of the Otis tests, forms *A* and *B*, verbal analogies, and artificial language. The average correlation of these four scores with the gain scores is negligible with the exception of the multiple instruction concellation test, in which case it is +.381. Perhaps equally interesting is the fact that none of the gain scores shows a significant loading with Factor V, which is regarded as a speed factor. Not even in the speed of making gates test were the gain scores correlated with the speed factor.

4. The factor loadings of the gain scores depend largely upon the factor loadings of the initial and final scores. Gain scores usually correlate highly with final scores, whereas their correlation with initial scores seems to fluctuate, widely it is true, about zero or a small negative value. To a certain extent, undoubtedly, the higher correlation of gain scores with final than with initial scores is due to errors of measurement, the errors in the final scores being added to, and those in the initial scores being subtracted from, the gain scores. When the reliability of initial and final scores is as high as in the present study, however, the effect of errors of measurement cannot be a major factor. The reason why gain scores correlate higher with final than with initial scores is simply that the formula for a gain score is, *plus* final score *minus* initial score. Consequently, gain scores fluctuate directly with final scores but inversely with initial scores. In view of this fact, when initial and final scores differ considerably in their loading with a given factor, one would expect to find the loading of the gain score follow that shown by the final score. In such a case, if an individual possessed in high degree an ability which was important for initial score but not important for final score, the possession of that ability to a high degree would not tend to result in a high gain score. On the other hand, a high degree of an ability entering more importantly into final than into initial score would almost guarantee a high gain score.

The results of the factor analysis harmonize well with the preceding considerations. The seven instances of most marked increase in final

over initial loading, all show significant gain-score loadings. For example, horizontal adding in the case of Factor IV shows an initial loading of +.239 and a final loading of +.727, and the gain-score loading is +.752. On the other hand, a high initial loading, particularly in the absence of an equally high final loading, tends to result in a low gain-score loading. There are ten instances, representing every one of the seven practice tests, in which the initial loading is +.296 or higher and the final loading is lower than the initial one. In all ten instances, the gain-score loadings are insignificant. For example, horizontal adding shows in the case of Factor VI an initial loading of +.448 and a final loading of +.270, and the gain-score loading is +.057. A more striking case is afforded by anagrams. This test shows an initial loading with Factor VI of +.945 and a final loading of +.744 and the gain-score loading is −.066. The interesting conclusion then appears to be clearly established, that, in the case of none of the seven tests here used, does the amount possessed of an initially important ability have any bearing upon the change with practice in an individual's standard score. The possession of such an ability to a high degree creates no likelihood of a greater than average or smaller than average gain. Even the possession of a high degree of an ability which shows a high final loading does not necessarily result in a high gain score. It will do so only providing the ability in question is less important in determining initial score than in determining final score.

5. Although no factor common to all or even a majority of the gain scores was discovered, several factors show loadings with the gain scores of more than one test. Factor II correlates +.497 with improvement in substitution and +.907 with improvement in Philip's cancellation; and Factor IX correlates +.405 with improvement in substitution and +.658 with improvement in speed of making gates. And Factor IV shows a high correlation with improvement in horizontal addition (+.752) and one which may not be negligible (+.378) with improvement in anagrams. It seems probable that if two tests have a common factor in their gain scores, practice in one would result in transfer to the other. Since, however, a high gain score loading appears always to be accompanied by a high final score loading, it is not likely that two tests will both show a high gain score loading with the same factor, unless this factor is also an important determinant of final score in both tests, and, further, unless it is a more important determinant of final than of initial score. It seems unlikely, then, that valid predictions as regards transference can be made by a factor analysis

of correlations of scores on tests given but once, or tests from which only a single score is used. Thus, for example, there is no reason for predicting transference from practice in Philip's cancellation to speed of making gates. While both these tests show a significant *initial* correlation with Factor V (+.684 and +.598), they do not both show a high final correlation with this or any other factor. Naturally, therefore, in view of considerations which have been outlined above, their gain scores show no sizeable correlation in common with any factor. On the other hand, transference might well be expected from practice in Philip's cancellation to substitution. This expectation would be based *not* on a high common initial loading (which, incidentally, they have with Factor V) but on the fact that *final* scores in both tests show a high loading with Factor II, and the further fact that this factor is altogether inconsequential in the initial scores of both tests. On account of the marked increase with practice in the loadings with Factor II, both tests also show marked correlation in their gain scores with this factor. This appears to be the state of affairs which would lead to transference of training. It would seem reasonably safe, then, to predict that in a group of subjects similar to the one here used, practice in Philip's cancellation would show transference to letter-digit substitution but not to speed of making gates.