A STUDY OF THE EFFECT OF LENGTH, PRACTICE, AND DEGREE OF LEARNING UPON THE RELIABILITY OF NONSENSE SYLLABLE SCORES

A THESIS

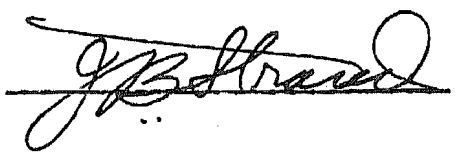
SUBMITTED TO THE DEPARTMENT OF PSYCHOLOGY AND THE GRADUATE COUNCIL OF THE KANSAS STATE TEACHERS COLLEGE OF EMPORIA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

By

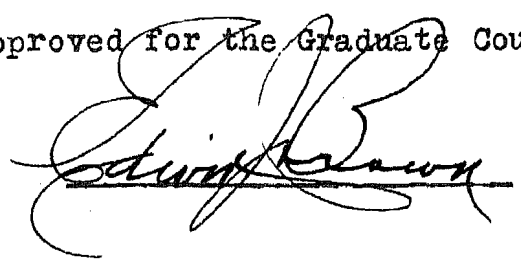
ADIN F. LEHMAN

July 1932

Approved for the Major Department



Approved for the Graduate Council



A C K N O W L E D G M E N T

To Doctor James B. Stroud, Chairman of the Department of Psychology, of the Kansas State Teachers College of Emporia, Kansas, who suggested the subject, revealed its possibilities, and directed the development of this thesis, the writer is especially indebted.

To the thirty-six fellow students who, as subjects, aided in the compilation of the necessary data by their voluntary contribution of both time and patience in learning the nonsense syllables, the writer gratefully expresses his appreciation.

TABLE OF CONTENTS

L I S T O F T A B L E S

I N T R O D U C T I O N

Statement of the Problem: The aim of this study is to determine the effect of length, practice, and degree of learning upon the reliability of nonsense syllable scores. This experiment made a simultaneous attack upon three distinct phases of the problem: (1) the effect of length of list upon the reliability of nonsense syllable scores; (2) the effect of practice upon the reliability of nonsense syllable scores; and (3) the effect of the degree of learning upon the reliability of nonsense syllable scores.

Historical Summary. Since much of the systematic work upon the subject of reliability of learning scores has been done in connection with the maze and to a lesser extent with the problem box, it is deemed advisable to review this work with special attention given to method.

Stone[1] in determining the reliability of time scores in the solving of the problem box and the maze by rats correlated the sums of time scores for any series of odd-numbered trials with the sums of time scores for even-numbered trials. His coefficients of reliability for the learning of rats to escape the maze from 1 to 10 trials was from .51±.09 to .68±.03; 11

---

1 Heron, William T. "Individual Differences in Ability versus Chance in Learning Stylus Mazes"; in COMPARATIVE PSYCHOLOGY MONOGRAMS, Vol. 2.

to 20 trials, .59±.04 to .67±.05; 6 to 20 trials, .49±.09 to .76±.02; and from 1 to 20 trials, .35±.12 to .53±.04.

Heron[2] in his maze experiments with animals in 1922 set up a criteron of learning and computed his coefficient of reliability on the total number of trials necessary to satisfy this criteron. Stone states that this method would produce higher coefficient correlations of reliability except when the total trials and time are highly correlated.

Webb[3] in a study of rats learning the maze correlated the learning scores with relearning scores (relearned 30 days after learning). He obtained only 16 positive correlations out of 30. He concluded that the majority of correlations was too small to be significant.

Hunter[4] in the experiment with rats learning a maze found a correlation of only .31 between scores on left and right turns. Human subjects in the same experiment by the same method of computation, showed a correlation of .80.

Heron and Hunter[5] in studying maze learning ability of rats used three different methods in determining reliability coefficients: (1) odd trials versus even trials, (2) the sum of the first six trials versus another six trials sixty days later, and

---

2 Stone, C. P. "The Age Factor in Animal Learning: 1. Rats in the Problem Box and the Maze"; in GENETIC PSYCHOLOGY MONOGRAPHS, Vol. 5, No. 1, p. 15.

3 Heron, William T. "Individual Differences in Ability versus Chance in Learning Stylus Mazes"; in COMPARATIVE PSYCHOLOGY MONOGRAMS, Vol. 2.

4 Ibid.

5 Heron, William T. and Hunter, Walter S. "The Reliability of the Inclined Plane Problem Box as a Method of Measuring the Learning Ability of the Rat"; in JOURNAL OF COMPARATIVE PSYCHOLOGY, 1922, Vol. 1.

(3) the sum of the first six trials in the problem box and the first six trials in a maze sixty days later. Each method produced a low correlation coefficient. A higher correlation will be obtained if the first two trials are discarded. By eliminating the first two trials a correlation of .50 was secured between odd and even trials. The second method (first 6 trials against first 6, 60 days later) resulted in a correlation of .51, but no correlation was found to exist between the maze and the problem box.

Hunter[6] using 31 human subjects with a simple pencil maze found foot rule correlations between succeeding tenths of from .20 to .85 with a median around .45 or .50. The reliability coefficient between succeeding tenths on a simple maze for rats will average around .38 to .40, according to Hunter. They are lower for the more complex mazes.

Stone[7] states that one method of obtaining reliability for time scores is to correlate the sums of time scores for any chosen series of odd-numbered trials with the sums of the time scores for the even-numbered trials. His coefficients of reliability for the learning of rats to escape the maze are as follows:

Trials	30 days old No. 80 Learning	10 days old No. 96 Learning	2 years old No. 28 Learning
1-10	.68 ± .03	.64 ± .03	.51 ± .09
11-20	.59 ± .04	.59 ± .03	.67 ± .05
6-20	.58 ± .04	.76 ± .02	.49 ± .09
1-20	.52 ± .05	.53 ± .04	.35 ± .12

---

6 Hunter, Walter S. "Correlation Studies with the Maze in Rats and Humans"; in COMPARATIVE PSYCHOLOGY MONOGRAMS, 1922, Vol. 1.

7 Stone, C. P. "The Age Factor in Animal Learning: 1. Rats in the Problem Box and the Maze"; GENETIC PSYCHOLOGY MONOGRAMS, Vol. 5, No. 1, pp. 1-131.

According to Stone, these coefficients of reliability for the time scores indicate sufficient reliability for group comparisons but not for comparisons of one individual with another.

From an historical point of view, this problem has received but very little experimental treatment. There have been many studies in memory of prose, poetry, numbers, and nonsense syllables, but the results regarding reliability are largely incidental.

Dr. Hermann Ebbinghaus who devised the so-called nonsense syllables was doubtless the first in the field and while his results were purely personal yet "Ebbinghaus's memory methods are today standard procedures in the psychological laboratory and his main results may be accepted substantially as he left them".[8]

Woodrow[9] reported the largest set of reliability coefficients in connection with his study of transfer. He used end-tests similar in form but different in content. He found the following coefficients: (1) rote poetry, .67; (2) rote prose, .49; (3) facts, .48; (4) historical dates, .60; (5) Turkish-English vocabulary, .70; (6) auditory memory span for consonants, .55.

Lemmon[10] by correlating repeated trials or alternate items reports coefficients of .60 for logical memory, .85 for auditory paired associates, .94 for visual paired associates, and .91

---

8 Garrett, Henry E. Greater Experiments in Psychology, pp.76-77.
9 McGeoch, John A. "Memory"; in PSYCHOLOGICAL BULLETIN, 1928, Vol. 25, p. 525.
10 Ibid., p. 525.

for Turkish-English substitution.

McGeoch[11] states that while there have been no systematic investigations on the reliability of memory experiments, such as was made by Hunter on mazes, yet there have been reliability coefficients made and reported occasionally.

The educational conclusions of the correlation coefficients will depend primarily upon the personal experience of the individual making the interpretation. Rugg states:[12]

> The experience of the present writer in examining many correlation tables has led him to regard correlations as "negligible" or "indifferent" when "r" is less than .15 to .20; as being "present but low" when "r" ranges from .15 or .20 to .35 or .40; as being "markedly present" or "marked" when "r" ranges from .35 or .40 to .50 or .60; as being "high" when it is above .60 or .70. With the present limitations on educational testing few correlations in testing will run above .70, and it is safe to regard this as a very high coefficient.

Purpose. Since the time of Ebbinghaus the nonsense syllable has been in general use in memory experiments in practically all psychological laboratories. It seems pertinent, therefore, to undertake a systematic study of nonsense syllables with a view to determining their reliability in general with special attention to some of the factors influencing their reliability.

One finds that varying lengths of list have been used by investigators. The length of list has usually been chosen to suit the convenience of the experimenter. It seems highly desirable, therefore, to undertake a study of the relation of length of list to the reliability of the scores.

---

11 McGeoch, John A. "Memory"; in PSYCHOLOGICAL BULLETIN, 1928, Vol. 25, pp. 513-49.

12 Rugg, Harold O., Statistical Methods Applied to Education, p. 256.

Subjects frequently come to the laboratory to participate in experiments with varying degrees of practice in learning nonsense syllables. It seems desirable, therefore, to know the effect of practice upon the reliability of the scores. Furthermore, should it be found possible to increase the reliability of scores by practice such practice would be desirable before undertaking an experiment. A portion of this study is devoted to this question.

The final aim of this study is to determine the relation of degree of learning required to the reliability of the scores. The reliability of the scores is determined, in this study, for the following degrees of learning: (1) right but two, (2) right but one, (3) all right, (4) right for two consecutive trials, and (5) right for three consecutive trials.

PROCEDURE

Materials. The materials for this experiment consisted of nonsense syllables of three letter type, a vowel enclosed by two consonants. They were written in capitals and presented visually to the subject in a single vertical column. The complete list is given in the appendix. (The first syllable in each list was used as a cue. It was not learned by the subject.)

Subjects. Thirty-six students (12 men and 24 women) selected at random in the Kansas State Teachers College of Emporia, Kansas, of graduate level or of senior college ranking participated in this experiment.

Method. The syllables were presented one at a time by the memory drum method. They were presented to subjects at intervals of about two seconds each. These they spelled out in audible recitation. The material was recalled by the method of anticipation. A maximum of three seconds per syllable was allowed for recall. If a subject anticipated a syllable in less than three seconds, he was immediately shown the next syllable. Presentation and recall were alternated until two and three successive and correct interpretations of each list were made.

Controls. The schedule for the learning of the six lists

of nonsense syllables was held constant, insofar, as possible both as to place and the time of meeting, for six consecutive days. All experimenting was conducted solely by the writer. In an experiment of this kind uniform difficulty of lists of the same length is imperative. The syllables for the various lists were selected at random. Furthermore, the syllables were varied systematically from list to list. Practice effects were controlled by a counter-balanced order. For example, subject 1 learned the lists (according to length) as follows: 6, 6, 12, 12, 18, and 18; subject 2 learned them in order 12, 12, 18, 18, 6, and 6; and subject 3 learned them in the order 18, 18, 6, 6, 12, and 12. The next three subjects would follow the same procedure and so on. Thus, of the thirty-six subjects, twelve learned the lists in the order 6, 12, and 18; twelve learned them in the order 12, 18, and 6; and twelve in the order 18, 6, and 12. The type and character of the subjects and their full knowledge of the importance of the results contributed greatly to the accuracy of the data. The schedule will be found in the appendix.

R E S U L T S AND D I S C U S S I O N

(1) The Effect of Length on the Reliability of Nonsense Syllable Scores

Table I shows the correlation between the number of trials for learning the similar lengths of nonsense syllables for the thirty-six subjects. The reader will bear in mind that lists of the same length were learned on consecutive days. Thus the correlations of table I represents the correspondence between the number of trials required to learn two lists of 6, 12, and 18 syllables, respectively.

TABLE I

CORRELATION COEFFICIENTS BETWEEN THE NUMBER OF TRIALS REQUIRED TO LEARN EACH OF THE THREE LENGTHS OF NONSENSE SYLLABLES

Lengths	"r"	P.E.
Six	.253	.1024
Twelve	.673	.0616
Eighteen	.483	.0800

Read table thus: The correlation between the successive learnings of the two groups of six nonsense syllables on trials is .238 ± .1024.

The correlation between the two lists of six syllables is positive but low. It is approximately 2.5 times its P.E. The true correlation is indicated statistically to lie between .151

and .355. The coefficient for the twelve syllables is .673 with a probable range[1] of from .611 to .735. The coefficient for the eighteen syllables is .483 with a probable range of from .403 to .563.

Table II shows the correlation between the time required for learning the nonsense syllables of similar lengths of material for the thirty-six subjects.

TABLE II

CORRELATION COEFFICIENTS BETWEEN THE TIME REQUIRED TO LEARN EACH OF THE THREE LENGTHS OF NONSENSE SYLLABLES

Lengths	"r"	P.E.
Six	.606	.0683
Twelve	.725	.0544
Eighteen	.422	.0896

Read table thus: The correlation in time between the learning of the two groups of six nonsense syllables is .606 ± .0683.

The results of table II show a closer agreement between time scores for each of the three lengths of material than that between trial scores, which seems to indicate that time is a more reliable criteron of learning. It is somewhat significant that the list of twelve syllables still remains the most reliable as to length in respect to time with a quite high correlation

---

1 The term range in this case refers to the fact that subsequent correlations of this nature would be expected to fall within the range indicated.

coefficient of .725 and a range of from .671 to .779. The list of six syllables produced a correlation coefficient of .606 with a range of from .538 to .674 while the list of eighteen syllables showed a correlation coefficient of .422 with a range of from .332 to .512. Although in the effect of length on trial scores the list of eighteen syllables ranked second, in the effect of length on time scores the six syllables and the eighteen syllables rank second and third, respectively.

There is no precedent in the literature as to the most valid procedure of determining the reliability of learning scores as there is in the field of mental testing. Consequently the writer is resorting to several methods. It would seem that the most clearly indicated method is that followed above, in which the scores on two different lists of the same length are correlated. A second attempt to study the reliability of nonsense syllable scores and the relation of their reliability to length of list is that of correlating the average number of correct recalls on odd and even trials.

The correlation coefficients between the average number of correct recalls of syllables on the even trials and those of the odd trials is shown in table III. There are seventy-two cases in this correlation, due to the fact that each of the thirty-six subjects learned two lists of six, twelve, and eighteen syllables.

TABLE III

CORRELATION COEFFICIENTS BETWEEN THE AVERAGE CORRECT RECALLS ON THE EVEN TRIALS VERSUS THE ODD TRIALS FOR EACH OF THE THREE LENGTHS OF NONSENSE SYLLABLES

Lengths	"r"	P.E.
Six	.568	.0516
Twelve	.342	.0734
Eighteen	.712	.0411

Read table thus: The correlation between the average number of correct recalls between the even trials and the odd trials for the six syllables is .568±.0516.

The magnitude of the correlations obtained by this procedure are in general agreement with those obtained by the method of correlating the scores on one list with those on a comparable list. The correlation coefficient on the average number of correct recalls between the even and odd trials for the eighteen syllables was .712 with a range of from .671 to .753. Thus this group ranks first. The list of six syllables ranks second with a correlation coefficient of .568 and a range of from .516 to .620. Contrary to the results as shown in tables I and II in which the list of twelve syllables ranks first, this group drops to third place with a correlation coefficient of .342 and a range of from .269 to .415.

A further correlational procedure has been followed in connection with this problem, namely, that of correlating the num-

ber of trials required to learn the odd-numbered items (syllables) with the number of trials required to learn the even-numbered items. Since each subject learned two lists of each length, a total of seventy-two cases is included in each correlation. Table IV shows the results.

TABLE IV

CORRELATION COEFFICIENTS BETWEEN THE NUMBER OF TRIALS REQUIRED TO LEARN THE ODD-AND EVEN-NUMBERED SYLLABLES FOR EACH OF THE THREE LENGTHS

Lengths	"r"	P.E.	Prediction for whole test "r"
Six	.709	.0411	.830 ± .0224
Twelve	.925	.0153	.961 ± .0079
Eighteen	.939	.0079	.968 ± .0079

Read table thus: The correlation between the number of trials required to learn the odd-and even-numbered syllables for the lists of six is .709 ± .0411 with a prediction for the whole test of .830 ± .0224.

The list of eighteen syllables retains first place in this procedure with a correlation coefficient of .939 and a range of from .931 to .945 and with a prediction correlation coefficient of .968 ± .0079 if the test were doubled. The list of twelve syllables ranks a very close second with a correlation coefficient of .925 and a range of from .91 to .94 and a prediction range for the whole test of .961 ± .0079. The list of six syllables shows a high correlation even though in third place. This correlation is .709 with a range of from .668 to .750 and

the prediction for the whole test is .830 .0224.

By this criteron (odd-numbered versus even-numbered items), the reliability coefficients are markedly high. Likewise, the trend of the results is different. Here the six syllable lists suffer in comparison with the larger lists in reliability.

Summing up the results thus far obtained, we find in the first place that the reliability coefficients, except in the case of the last method used, are comparatively low. In the second case, marked variability in magnitude of the correlations appears for the different methods used for the different lengths of material. Even in the last two procedures where 72 cases were used marked variations occurred in the relative magnitude of reliability coefficients for the different lengths. These observations are sufficient to raise considerable skepticism regarding the reliability of nonsense syllable scores in general. When the reliability coefficients are so unstable for 36 and even 72 cases, there is considerable ground for doubt regarding much of their present usage in experimental procedure where the number of cases is usually smaller than the number used in this study.

Regarding the question of the relation of length to the reliability of the scores, it would seem that the scores themselves are too unreliable to warrant any very definite conclusions. However, inspection of the foregoing data reveals the fact that in the four sets of correlations, the reliability coefficients for the 12-and 18-syllable lists are higher than those of the 6-syllable lists in three of the four comparisons. In two of the four comparisons the reliability coefficients of the 18-syllable lists are higher than those of the 12-syllable lists.

Comparing the average reliability coefficients for the three lengths of material by the three methods, the following results are obtained: the average reliability coefficient for the 6-syllable list is .50; the average for the 12-syllable list, .66; and that for the 18-syllable list, .71. These comparisons, while not strongly insisted upon, seem to indicate that within the limits of this experiment, reliability of nonsense syllable scores increases with length of list.

(2) The Effect of Practice on the Reliability of Nonsense Syllable Scores

The effect of practice upon the reliability of nonsense syllable scores for the various lengths of syllables in their various positions of the schedule is shown in table V. Twelve cases were used in each instance because of the fact that twelve subjects learned two lists each of the three lengths of materials in each position with respect to practice.

TABLE V

CORRELATION COEFFICIENTS FOR TRIALS BETWEEN EACH COMPARABLE LENGTH OF MATERIAL IN EACH POSITION

Lengths and Days	Rho	P.E.
Six		
1st Day vs 2nd Day	.356	.1267
3rd Day vs 4th Day	.493	.1131
5th Day vs 6th Day	.580	.0965
Twelve		
1st Day vs 2nd Day	.375	.1267
3rd Day vs 4th Day	.958	.0147
5th Day vs 6th Day	.632	.0871
Eighteen		
1st Day vs 2nd Day	.716	.0660
3rd Day vs 4th Day	.340	.1267
5th Day vs 6th Day	.587	.0965

Read table thus: The correlations on trials between the lists of six syllables for the twelve subjects learning them on the same day is .356 ± .1267.

The reader will observe that for the list of six syllables, there is a tendency for the reliability coefficients to improve with practice on successive days. Between the scores on the two six syllable lists learned on the first and second days, a coefficient of .356 was found; a coefficient of .493 was found between the scores on two six syllable lists learned on the third and fourth days; while between the fifth and sixth days, the correlation coefficient was found to be .580.

The correlation coefficients between the scores on the twelve syllable lists for the respective practice days are as follows: 1st and 2nd days, .375 ± .1267; 3rd and 4th days, .958 ± .0147; 5th and 6th days, .632 ± .0871. Some of this fluctuation is doubtless due to chance.

The correlation coefficients between the scores on the eighteen syllable lists are as follows: 1st and 2nd days, .716 ± .0660; 3rd and 4th days, .340 ± .1267; 5th and 6th days, .587 ± .0965. In the case of the eighteen syllable lists, there is very little tendency for reliability to vary with practice.

Table VI is identical with table V in every respect, except that it is computed on the basis of the time required for learning rather than the number of trials. All other conditions are the same. Twelve cases are included.

TABLE VI

CORRELATION COEFFICIENTS FOR TIME BETWEEN EACH COMPARABLE LENGTH OF MATERIAL IN EACH POSITION

Length and Days	Rho	P.E.
Six		
1st Day vs 2nd Day	.711	.0660
3rd Day vs 4th Day	.535	.0965
5th Day vs 6th Day	.790	.0543
Twelve		
1st Day vs 2nd Day	.636	.0871
3rd Day vs 4th Day	.564	.0965
5th Day vs 6th Day	.552	.0965
Eighteen		
1st Day vs 2nd Day	.582	.0965
3rd Day vs 4th Day	.139	.1493
5th Day vs 6th Day	.697	.0769

Read table thus: The correlation on time between the lists of six syllables for the twelve subjects learning them on the same day is .711 ± .0660.

The inconsistency of these coefficients of correlation is quite noticeable. In the list of six, the correlation coefficient drops from .711 ± .0660 to .535 ± .0965 and then rises to .790 ± .0543. In the list of twelve, the correlations make a more steady decline from .636 ± .0871 between the first two days to .564 ± .0965 for the third versus the fourth day and then to .552 ± .0965 for the fifth versus the sixth day. The list of eighteen declines from .582 ± .0965 for the first two days to a negligible coefficient of .139 ± .1493 for the third day versus the fourth day and then rises to .697 ± .0769 for the fifth

versus the sixth day. These variations are about what would be expected upon the basis of chance. There seems to be no clear tendency for reliability to vary with practice when reliability is computed by this method.

The effect of practice on successive days for the entire lists of material, regardless of length, is shown in table VII. This correlation includes both the trials and the time for the thirty-six subjects. The attention of the reader is called to the fact that the coefficients in table VII are not true reliability coefficients. There is a constant spurious factor running through them, in that the scores for the various lengths were plotted upon the same correlation chart. However, inasmuch as, this spurious factor is constant for the successive days it does not invalidate the comparison of the coefficients with reference to practice effect.

TABLE VII

CORRELATION COEFFICIENTS FOR BOTH TRIALS AND TIME ON SUCCESSIVE DAYS FOR THE ENTIRE MATERIAL, REGARDLESS OF LENGTH

Days	Trials		Time (Min.)	
	"r"	P.E.	"r"	P.E.
1st Day vs 2nd Day	.809	.0384	.882	.0203
3rd Day vs 4th Day	.749	.0467	.725	.0467
5th Day vs 6th Day	.846	.0296	.912	.0203

Read table thus: The correlation coefficient between the entire material for the first and second days is .809 ± .0384 on trials and .882 ± .0203 on time.

The difference in these correlation coefficients seems to indicate that practice had very little effect on successive days. The coefficients between the first two days which are .809±.0384 for trials and .882±.0203 for time decreases for the third and fourth days to .749±.0467 for trials and .725±.0467 for time and then rises for the fifth and sixth days to .846±.0296 for trials and .912±.0203 for time.

A further attempt was made to show the effect of practice upon reliability by correlating the average correct recalls on the odd trials with those on the even trials for the entire material (regardless of length) for each of the six days. The same spurious factor exists here as that to which attention was called in connection with table VII. This is shown in table VIII. The thirty-six subjects are included.

TABLE VIII

CORRELATION COEFFICIENTS BETWEEN THE AVERAGE CORRECT RECALLS ON THE ODD VS EVEN TRIALS FOR THE ENTIRE MATERIAL FOR EACH OF THE SIX DAYS

Days	"r"	P.E.
1st Day	.918	.0203
2nd Day	.919	.0203
3rd Day	.912	.0203
4th Day	.907	.0203
5th Day	.937	.0104
6th Day	.858	.0296

Read table thus: The correlation between the average correct recalls of odd and even trials for entire material for the 1st day is .918±.0203.

These data are in agreement with those presented previously in that they show that practice as defined in this experiment has very little effect on the reliability of nonsense syllable scores. The coefficient for the fifth day was .937 while the coefficient for the sixth day was .858. Considering the six practice days, these comprise the highest and lowest correlation coefficients, making the range of all days from .858 to .937, an insignificant difference of only .079.

The effect of practice on the various lengths of material is also shown by correlations in table IX. These are correlations between the average correct recalls on the odd-and even-numbered trials for the 1st, 3rd, and 5th days. Twelve cases were used in each correlation because of the fact that twelve subjects learned similar lengths of material on the same days.

TABLE IX

CORRELATION COEFFICIENTS BETWEEN THE AVERAGE CORRECT RECALLS ON THE ODD VS EVEN TRIALS FOR SIMILAR LENGTHS OF MATERIAL FOR THREE DAYS

Lengths	1st Day		3rd Day		5th Day	
	Rho	P.E.	Rho	P.E.	Rho	P.E.
Six	.622	.0871	.569	.0965	.573	.0965
Twelve	.005	.1508	.211	.1448	.341	.1267
Eighteen	.496	.1131	.608	.0871	.786	.0543

Read table thus: The correlation coefficients between trials for the six syllables for the first day is .622 ±.0871; for the third day, .569±.0965; and for the fifth day, .573±.0965.

The results given in table IX also indicate that reliability is unaffected by practice. The coefficients for the six

syllables decline from .622±.0871 for the first day to .573± .0965 on the fifth day, making a range of from .535 to .670. The coefficients for the twelve syllables increase from .005± .1508 on the first day to .341±.1267 on the fifth day, making a range of -.1458 to .468. The list of eighteen syllables has a coefficient of .496±.1131 for the first day and .786±.0543 for the fifth day, making a range of from .383 to .840.

The correlation coefficients between the number of trials required to learn the odd-numbered and the even-numbered syllables for each of the six days, regardless of the length of the material, together with the prediction (by the Spearman-Brown formula) for the whole test, is shown in table X. This table includes thirty-six cases, representing the thirty-six subjects. Again attention is called to the fact that a constant spurious factor, namely, that of varying lengths of material, runs throughout these correlations.

TABLE X

CORRELATION COEFFICIENTS BETWEEN THE NUMBER OF TRIALS REQUIRED TO LEARN THE ODD AND THE EVEN-NUMBERED SYLLABLES FOR THE SIX DAYS

Days	"r"	P.E.	Prediction for whole test "r"
1st Day	.961	.0104	.980 ± .0104
2nd Day	.946	.0104	.972 ± .0104
3rd Day	.967	.0104	.983 ± .0104
4th Day	.968	.0104	.983 ± .0104
5th Day	.957	.0104	.978 ± .0104
6th Day	.955	.0104	.977 ± .0104

Read table thus: The correlation coefficient between the number of trials required to learn the odd-numbered and the even-numbered syllables, regardless of length of material for the first day is .961 ± .0104 with a prediction of .980 ± .0104 if the test were doubled.

These coefficients are practically constant for all practice days. The highest occurs for the fourth day with a coefficient of .968 ± .0104 and a prediction coefficient for the whole test of .983 ± .0104. The lowest occurs for the second day with a coefficient of .946 ± .0104 and a prediction coefficient for the whole test of .972 ± .0104. The results of these correlations show neither a decrease nor an increase of sufficient amounts to warrant attaching any degree of significance to the effect of practice upon reliability.

The various methods used in studying the effect of practice upon the reliability of nonsense syllable scores agree with re-

markable uniformity in showing that within the limits of this experiment, practice is non-important.

(3) The Effect of the Degree of Learning on the Reliability of Nonsense Syllable Scores

The relation between degree of learning and the reliability of the scores is studied by correlational procedure. This is done by correlating the trials on comparable lists in which all syllables were right but two, all were right but one, all were right, and all were right twice. Each of the thirty-six subjects learned three lists of nonsense syllables, of varying length as listed above, on three alternate days, making a total of one hundred and eight cases. These scores were correlated with the scores on three lists of the same lengths for the other three days. The reader will note that there is a constant spurious factor of length which operates to raise all the coefficients. But since it is constant it would operate to the same extent in all comparisons. Thus any differences in the magnitude of the correlations would be due to the effect of degree of learning or to chance. These results are shown in table XI.

TABLE XI

CORRELATION COEFFICIENTS ON TRIALS SHOWING THE DEGREE OF LEARNING FOR THE ENTIRE MATERIAL

Degree	"r"	P.E.
Right but 2	.831	.0187
Right but 1	.801	.0242
All right	.788	.0242
Right twice	.789	.0242

Read table thus: The correlation between the entire list of six, twelve, and eighteen syllables for three days in which all were right but 2 with the second list for the other three days is .831 ± .0187.

The comparative closeness of all the correlations indicates that one degree of learning is about as good a criteron as the other. The correlation coefficient for "those right but two" was .831 ± .0187; for "those right but one", .801 ± .0242; for "those right", .788 ± .0242; and for "those right twice", .789 ± .0242. In no instance is any great degree of variability revealed, the range from the highest to the lowest being from .788 to .831. These variations are well within the realm of chance.

The writer has purposely omitted the correlation of scores "right three times" due to the fact that only one subject out of the entire group of thirty-six failed to recall the syllables correctly for the third time after he had correctly recalled them twice and this person failed in only two instances.

Table XII is a correlation for the degree of learning in

much the same manner as table XI, except that the different lengths are considered, thus making only thirty-six cases in each correlation. It seemed desirable to make calculations similar to those presented in table XI for each length of material. It is possible that the relation of the reliability of the scores to degree of learning might vary with the different lengths of material. The results are given in table XII.

TABLE XII

CORRELATION COEFFICIENTS ON TRIALS SHOWING THE DEGREE OF LEARNING FOR THE DIFFERENT LENGTHS OF MATERIAL

Degree	"r"	P.E.
Six		
Right but two	.241	.1024
Right but one	.402	.0896
All right	.218	.1024
Right twice	.238	.1024
Twelve		
Right but two	.658	.0616
Right but one	.652	.0616
All right	.781	.0384
Right twice	.673	.0616
Eighteen		
Right but two	.533	.0800
Right but one	.505	.0800
All right	.441	.0896
Right twice	.483	.0800

Read table thus: The correlation between the lists of six syllables in which all were right but two is .241 ± .1024.

Examination of these data fails to show any tendency for reliability to vary with the degree of learning for any length

of material used. The fluctuations are within the realm of chance.

A supplementary attempt to study the reliability of nonsense syllable scores is shown in tables XIII and XIV. The averages in both time and trials in the memorizing of the various lengths of nonsense syllables are shown in table XIII.

The reader will bear in mind that each subject learned two six syllable lists, two twelve syllable lists, and two lists of eighteen syllables. It seems worthwhile to compare the average mean scores between the two comparable lists both as to time and trials. Any large discrepancy between the lists of the same length would be construed as the result of unreliability of the scores or to practice since the lists, by the method stated above, have been equated as to difficulty.

TABLE XIII

AVERAGES OF THE TIME AND TRIALS FOR THE LEARNING OF THE PAIRED LENGTHS OF MATERIAL

Lengths	First Group Time(Min.)	Trials	Second Group Time(Min.)	Trials
Six	3.19	4.31	2.63	3.94
Twelve	14.82	9.14	12.38	8.67
Eighteen	31.78	13.06	24.65	11.67

Read table thus: It took an average of 3.19 minutes for each of the 36 subjects to learn the 6 nonsense with an average of 4.31 trials but on the second list, it required an average time of 2.63 minutes and an average of 3.94 trials.

It will be observed in table XIII that the averages in both the length of time and the number of trials required for learning the various lists improved between the paired groups. The

time and trials both seemed to improve at a constant rate. The length of time for the six syllables decreased .56 minutes; for the twelve syllables, 2.44 minutes; and for the eighteen syllables, 7.13 minutes. The average number of trials decreased .37 for the six syllables, .47 for the twelve syllables, and 1.39 for the eighteen syllables.

If the disparity between the scores of comparable lists were due to chance (unreliability) it ought to be of a random nature. The fact that the discrepancies are all in the same direction, the scores for learning the second lists being lower by all comparisons than those of the first lists, and that the differences are directly proportional to the length of lists, indicates that but for practice the mean scores for comparable lists would be almost identical.

It seemed advisable to see to what extent the same tendency would hold for smaller groups. Consequently, the thirty-six subjects were divided, by chance selection, into three groups of twelve each. The median time and trial scores required to learn the first list are compared with those required to learn a second list of the same length. This is shown in table XIV.

TABLE XIV

THE PAIRED MEDIANS FOR THE TWO LISTS OF EACH LENGTH OF NONSENSE SYLLABLES FOR THE CHANCE SELECTION OF THE THREE GROUPS

Length of Material	Median First Group Time (Min.)	Trials	Median Second Group Time (Min.)	Trials
Eighteen	31.17	13.5	24.40	13.0
Eighteen	31.32	13.0	22.10	12.0
Eighteen	31.67	12.0	25.46	11.0
Twelve	12.88	9.0	10.98	9.0
Twelve	11.42	8.5	9.69	9.0
Twelve	12.50	8.0	13.79	7.5
Six	2.88	5.0	2.59	4.5
Six	3.59	5.5	2.46	5.0
Six	4.17	5.0	3.88	4.5

Read table thus: The median score for the learning the list of eighteen nonsense syllables by a random selection of twelve subjects was 31.17 minutes with an average of 13.5 trials. The average time and trials for learning the second list of eighteen syllables were 24.40 minutes and 13 trials, respectively.

In the above comparisons there are only two cases which do not conform to the tendency expressed in the mean scores for the entire group. These are the second and third groups of the 12-syllable lists. This procedure argues for considerable reliability of nonsense syllable scores even with fairly small groups.

Following is a further attempt to study the general reliability of nonsense syllable scores. In table XV is given the percentage of subjects who made the same trial scores on two lists of the same length, learned on consecutive days, for each length of material.

TABLE XV

SHOWING A CONSISTENCY WITH WHICH SUBJECTS TENDED TO MAKE THE SAME SCORES ON TWO COMPARABLE LISTS OF NONSENSE SYLLABLES

Description	Eighteen Syllables percent	Twelve Syllables percent	Six Syllables percent
Identical scores	25	17	19
Difference of 1	14	33	47
Difference of 2	17	30	8
Difference of 3	17	5	14
Difference of 4	14	5	11
Difference of 5	5	5	0
Difference of 6	3	0	0
Difference of 7	3	0	0
Difference of 8	0	3	0
Difference of 12	3	0	0

Read table thus: Twenty-five percent of the subjects required the same number of trials in learning eighteen syllables of two lists, seventeen percent required the same number of trials for the twelve syllable list, and nineteen percent required the same number of trials in learning the two lists of six nonsense syllables.

These percentages show that there is a fairly consistent tendency for subjects to make the same or nearly the same scores on two lists of nonsense syllables of the same length. These data show that differences of two or three trials between two or more experimental conditions with a reasonably adequate number of subjects may be regarded with reasonable certainty as not

due to chance. According to table XV, the scores on one six syllable list do not differ from those on the other six syllable list by more than three trials in 88% of the cases, by more than two trials in 74% of the cases, and by more than one trial in 66% of the cases. For the twelve syllable lists, the scores on the two lists do not differ from each other by more than three trials in 87% of the cases, by more than two trials in 82% of the cases, and by more than one trial in 50% of the cases. For the eighteen syllable lists, the scores on two lists do not differ from each other by more than three trials in 73% of the cases, by more than two trials in 56% of the cases, and by more than one trial in 39% of the cases.

Further investigation shows that there was no case of a subject taking as many trials or as much time to learn a list of six syllables as he took to learn a list of twelve syllables. There are seventy-two cases in this comparison. Furthermore, there are only three cases out of the seventy-two which required as much time or as many trials to learn a list of twelve syllables as was required to learn a list of eighteen syllables.

C O N C L U S I O N S

Within the limits of this experiment, the following conclusions seem warranted:

1. There is a slight tendency for the reliability of nonsense syllable scores to increase with length.

2. Practice, while it has the obvious effect of reducing time and trial scores, has no apparent effect upon the reliability of the scores.

3. Five degrees of learning were employed in this experiment. The trial scores obtained by the various methods appear to be equally reliable.

4. The magnitude of the reliability coefficients obtained vary with the method employed. The method of correlating odd-against even-numbered items gave the highest coefficients, ranging from .95 to .97. The method of correlating the average scores on odd trials against those on even trials gave the next highest coefficients, ranging from .86 to .94. The method of correlating scores of two lists of syllables of the same length gave coefficients of about .50 with a wide range of scatter.

B I B L I O G R A P H Y

Garrett, Henry E. Greater Experiments in Psychology, The Century Company, New York, 1930. 337 pages.

Heron, William T. "Individual Differences in Ability versus Chance in Learning Stylus Mazes"; in COMPARATIVE PSYCHOLOGY MONOGRAMS, Vol. 2.

Heron, William T., and Hunter, Walter S. "The Reliability of the Inclined Plane Problem Box as a Method of Measuring the Learning Ability of the Rat"; in JOURNAL OF COMPARATIVE PSYCHOLOGY, 1922, Vol. 1.

Hunter, Walter S. "Correlation Studies with the Maze in Rats and Humans"; in COMPARATIVE PSYCHOLOGY MONOGRAMS, 1922, Vol. 1.

McGeoch, John A. "Memory"; in PSYCHOLOGICAL BULLETIN, 1928, Vol. 25, p. 525.

Rugg, Harold O. Statistical Methods Applied to Education, Houghton Mifflin Company, Boston, 1917. 410 pages.

Stone, A. P. "The Age Factor in Animal Learning: I. Rats in the Problem Box and the Maze"; in GENETIC PSYCHOLOGY MONOGRAMS, Vol. 5, No. 1, pp. 1-131.

A P P E N D I X

Materials. The following twelve lists comprise the series of nonsense syllables:

List (1)	List (7)	List (4)	List (10)	List (6)	List (12)
NEZ	LIR	TUH	CES	BOV	ZOB
GAH	TEZ	JIR	PIB	HIJ	NAJ
DUQ	QIH	KEC	ZOH	MEF	CEX
HIF	DIR	RIS	NIS	NAS	TUD
PEJ	KUH	CAZ	MOJ	GIQ	FAJ
BAV	SEJ	BUH	ZAF	XUR	DAX
LIJ	XOG	XAV	TIR	FIH	ZEN

List (2)	List (8)	List (5)	List (11)
ZOS	POB	JIH	VUM
CUH	CUG	DUR	CIB
GOC	TEQ	FEG	FAP
DIJ	FAH	ZAJ	NID
XEH	QAJ	BIH	JUF
VAF	VEH	DEG	QEM

List (3)	List (9)
GOX	BIR
BEF	FUB
XAB	CAH
NUR	KIR
DOK	NUV
FET	SEB

Schedule. The following schedule was followed in the learning of the nonsense syllables:

Group No.	1st Day Lists	2nd Day Lists	3rd Day Lists	4th Day Lists	5th Day Lists	6th Day Lists
I	1 2 3	7 8 9	4 5	10 11	6	12
II	5 6	11 12	1	7	2 3 4	8 9 10
III	2	8	3 4 5	9 10 11	6 1	12 7
IV	4 5 6	10 11 12	1 2	7 8	3	9
V	2 3	8 9	4	10	5 6 1	11 12 7
VI	5	11	6 1 2	12 7 8	3 4	9 10