A STUDY OF THE EFFECT OF LENGTH, PRATICE, AND

DEGREE OF LEARNING UPON THE RELIABILITY

OF NONSENSE SYLLABLE SCORES AND

RELATION OF LENGTH OF LIST

TO ITS DIFFICULTY

A THESIS

SUBMITTED TO THE DEPARTMENT OF

PSYCHOLOGY AND GRADUATE COUNCIL OF THE

KANSAS STATE TEACHERS COLLEGE OF EMPORIA

IN PARTIAL FULFILLMEN OF THE REQUIREMENTS

FOR THE DEGREE OF MASTER OF SCIENCE

BY

CELIA J. MC CUE

Approved for the Major Department

_____

Approved for the Graduate Council

_____

TABLE OF CONTENTS

LIST OF TABLES

INTRODUCTION

Statement of the Problem. The purpose of this study
is to make further contributions to the problems investigat-
ed by Lehman[1] namely: (1) the effect of the length of list
upon the reliability of nonsense syllable scores; (2) the
effect of pratice upon the reliability of nonsense syllable
scores; and (3) the effect of the degree of learning upon
the reliability of nonsense syllable scores. Inasmuch as the
two studies taken together include a comparatively large
number of cases, and since the data are admirably arrange
for the purpose, the relation of length of the lists to
difficulty is shown.

Historical Summary. Lehman's study[2] is the first one
dealing with the reliability of nonsense syllables. Very
little other work on reliability of learning measures has
been done with human subjects, especially in so-called
memory tasks. The reliability of the maze has received con-
siderable attention from the psychologist. This paper will
present a brief summary of the animal and human studies
bearing upon the reliability of learning measures together
with an account of the studies dealing with the effect of
length of task upon its difficulty.

---

1. Lehman, Adin F. A study of the Effect of Length,
   Pratice and Degree of Learning Upon
   the Reliability of Nonsense Syllable
   Scores. Published Kansas State
   Teachers College, Emporia, 1932

2. Ibid.

The work on reliability of learning material started
with Hunter and his students in 1922. Heron,[3] using the
inclined plane box and the Watson circular maze with rats,
got reliability coefficients generally between .30 and .40.
Somewhat similar results were obtained by Heron[4] in a later
study of the reliability of a stylus maze, and by Hunter[5]
in a study employing rats and human subjects in which mazes
of different complexity were used. Likewise, studies by
Hunter and Randolph, Liggett, Tolman and Davis have con-
firmed Heron's original findings regarding the reliability
of mazes.

Several later studies have shown reliability coefficients
much above those just cited. The reliability of the maze
scores, an instrument in psychological experimentation, was
found so unsatisfactory that considerable effort has been
made to improve this condition. The results are immediately
reflected in higher reliability coefficients. The results
of these studies follow.

Tryon[6] used 107 rats, training was given on his first
(mechanical)maze and on a second (hand operated) one. On

3. Leeper, Robert. "The Reliability and Validity of
        Maze Experiment With White Rats,"
        in GENETIC PSYCHOLOGY MONOGRAPHS,
        1932, p. 145.
4. Ibid. See number 5 in Bibliography
5. Ibid. See number 7 in Bibliography
6. Leeper, Robert. "Reliability and Validity of Maze
        Experiment with White Rats,"
        GENETIC PSYCHOLOGY MONOGRAPHY,1932,p.49

the first of the two mazes, animals of the same sex were given same quantity of food regardless of the fact that the age range was from three to eight months. On the second maze each animal was allowed to eat until he first turned away from the pan. High correlations were found between the two mazes. Errors from groups of three trials on the first maze were correlated with the errors of groups of three on the second maze(20 trials were given on each maze and the first and twentieth trials were dropped), the resulting raw coefficient ranged from .318 to .772 with the median coefficient .608.

Stone[7] received relatively high positive correlation when he computed the reliability of the time scores in solving problem box. All coefficients were above .35 and lower than .76. These correlations have a wide range but are quite significant. The coefficients were between odd and even trials. In his experiment with rats and the maze his correlations of odd against even trials were likewise between .35 and.76. This would indicate that both the problem box and the maze are more reliable measures than was indicated by the earlier work.

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

Stone and Nyswander have given the simple T maze the most reliable rating. This fact was brought out in connection with their study of the influence of age upon learning. They found reliability coefficients of .80 to .90, figured on eight groups of about twenty five rats each. Leeper[8] believes this may be due to systematic errors rather than such positive element in age. Nevertheless, the claims of Stone and Nyswander have been confirmed by Tolman and Nyswander.

Leeper[9] found that strong motivation seems to yield more reliable results than moderate motivation; that time scores in terms of trials to learn are less reliable than than error scores. He states:

> A comparison of these earlier experiments
> with more recent ones yielding higher co-
> efficients seems to indicate that the feature
> of the early experiments responsible for
> the low reliability coefficients were: (1)
> the fact that the maze were too simple and
> easy, (2) the lack, in most cases, of
> preliminary training to accustom the animals
> to the apparatus and handling to develop
> stronger motivation, (3) poor control of
> motivation, (4) the use, in some cases,
> of mazes with alleys of such unequal
> complexity that chance blundering into cer-
> tain alleys offered much greater hidrance
> to learning than blundering into others,
> and (5) the lack of means of preventing
> retracing, and (6) utilization of too few
> trials to furnish the data correlated.
> In some experiments, for instance, the

---

8. Leeper, Robert,   "Reliability and Validity of Maze
                     Experiments with White Rats," in
                     GENETIC PSYCHOLOGY MONOGRAPHS,
                     1932, p.15.

9. Ibid. p.145.

> correlated scores were errors for dif-
> ferent single trials; in other cases,
> groups of only three trials were cor-
> related.

As the writer earlier stated little work has been done
on reliability of memory scores. One of the most important
studies is that of Woodrow[10] made in connection with his
investigation of transfer. He used human subject, and gave
end tests that were similar in form but different in content.
He used different types of material, rote poetry, rote prose,
facts, historical dates, Turkish-English vocabulary, and
auditory memory span for constants. The highest coefficient
was in memorizing Turkish-English vocabulary which was .70,
the lowest was found for rote prose, which was .49.

Lemmon[11] in his study of memory reported highest re-
liability coefficients for the memorization of auditory
paired associations .94, which was followed by Turkish-
English substitution .91.

Lehman[12] in a companion study to the present investig-
ation found product moment correlation coefficient between
comparable tests of nonsense syllables to be, on an average,
in the neighborhood of .69. The other results of his study

---

10. Lehman, Adin F. A Study of Effect of Length, Pratice
    and Degree of Learning Upon Reliabil-
    ity of Nonsense Syllable Scores.
    Published Kansas State Teachers
    College, Emporia, 1932. p.4
11. Ibid.
12. Op. cit.

are summarized in his own words as follows:

1. There is a slight tendency for the reliability of nonsense syllable scores to increase with length.
2. Pratice, while it has the obvious effect of reducing time and trial scores, has no apparent effect upon the reliability of the scores.
3. Trial scores obtained by various methods appear to be equally reliable.
4. The magnitude of reliability coefficient obtained vary with method employed.

The reliability coefficients of studies on memory tasks have been quite low for precision instruments, especially for studies of individual differences. Perhaps time has honored the technique developed by Ebbinghouse more than is justified. Attention should be paid to the construction of more precise measures of learning for human subjects or to the improvement of existing technique.

Many investigators have reported data bearing upon the relation of length of material to its difficulty. Inasmuch as the present data are so well suited to this purpose the writer has sought to utilize these studies in an effort to throw additional light, if possible, upon this problem. Since this is only an incidental phase of the problem, an exhaustive historical summary is not required. However, the results of two or three typical studies will be cited for the purpose of acquainting the reader with the present status of the problem.

Thurston,[13] using Lyon's Brnet's and Henri's data, found

---

13. Thurston, L. L. "The Relation Between Learning Time and Length of Task" in PSYCHOLOGICAL REVIEW, 1930, 37.pp.44-53.

that learning time increases at the 3/2's power of the length of the list. That is, $T = C/K \; N \; 3/2$ in which T is the learning time, C and K are arbitary constants and N is the length of list. This curve was obtained from results in both nonsense syllables and digits. He also found that a negatively acclerated curve fit their data, when number of repetition was the pratice criterion. The equation was $R = C/K\sqrt{N}$ in which R is number of repetitions necessary to learn, C and K are arbitary constants and N is the length of lists. The curves also fit Ebbinghouse's[14] data. Another way of stating the results of their agreement would be to say that the longer the lists the more economical it is in terms of number of repetitions necessary to learn, but the less economical in terms of learning time.

Heron and Robinson[15] found that the number of repetitions required for complete learning increases rapidly, with early increase in length of material and more slowly with later ones. They also found that twelve syllable lists learned immediately after other lists are learned slightly more readily than those learned without a previous list. This positive transfer was found to be most marked when first lists were shorter.

14. Ebbinghouse, Hermann, Memory, A contribution to Experimental Psychology, Translated by Ruzer and Bussinius, Pub. Teachers College, Columbia, N. Y. 1913, pp.81-120

15. Robinson, E. S. & Heron, W. T. "Results of Variations in Length of Memorized Material" JOURNAL OF EXPERIMENTAL PSYCHOLOGY, Vol. 5, 1922. pp 428-449

Tomlinson[16]found in his study that there was a pro-
gressive increase in time as length of the series increas-
ed in memorizing rote material. He also found that in-
creasing the length of poetry does not have as great ef-
fect upon its difficulty as does the increasing the length
of list of nonsense syllables. His curve, in case of poetry,
does tend to show slight tendency for difficulty to increase
slower than length.

There have been many studies made in regard to length of
list and difficulty of rote material. They all tend to find
as length of series is increased its difficulty increases
at an increasing rate.

---

16. Tomlinson, Charles, Studies in Memory, Pub. Kansas
    State Teachers College, Emporia,
    1932 p.18

PROCEDURE

Materials. The materials used for this experiment consisted of three lettered nonsense syllables, a vowel enclosed by two consonants. The syllables were written in capital letters and presented visually to subjects in a single vertical column. (See appendix for entire list. The first syllable in each list was used as a cue word.)

Subjects. A group of forty students (33 women and 7 men) participated in this experiment. The subjects were selected at random and no attention was given to their educational accomplishments or their native ability to memorize. The group was composed of undergaruates as well as graduate students of Kansas State Teacher College, Emporia. It follows that there was a wide range in their ability to memorize the material.

Method. The syllables were presented one at a time by the memory drum method. The subjects spelled the syllables in auditory recitation. The material was recalled by the method of anticipation. They were presented in two second intervals, and three seconds were allowed for recall. If the subject recalled the syllable in less than three seconds, he was immediately allowed to recall the next one. Presentation and recall were alternated until three successive and correct interpretations of each lists were made.

Correlations and rhos were also made for the following degrees of learning (1) all right but two, (2) all right but one, (3) all right, and (4) all right twice in succession.

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 by the writer. The syllables were selected at random; they were varied systematically from list to list. Pratice effects were controlled by a counter balanced order. Subjects 1, 4, 7, 10, 13, 16, 19, 22, 25, 28, 31, 34, 37, and 40 learned the lists according to length as follows: 6, 6, 12, 12, 18, 18; subjects 2, 5, 8, 11, 14, 17, 20, 23, 26, 29, 32, 38, learned the lists in this order: 12, 12, 18, 18, 6, 6,; subjects 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, learned the lists in this order: 18, 18, 6, 6, 12, 12. (See Appendix for the schedule.)

RESULTS AND DISCUSSIONS

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

The correlations between trial scores of comparable lists of nonsense syllable for the three paired lengths are shown in Table I.

TABLE I

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

Lengths	"r"	P. E.
Six	.74	.01257
Twelve	.74	.01257
Eighteen	.84	.03144

Read table thus: The correlation between the sucessive learnings of two groups of six nonsense syllable on trials is .74±.01257.

The correlations in this experiment are higher than those obtained by Lehman[17] in his study. This may be due

17. Lehman, Adin F. A Study of The Effect of Length, Pratice, and Degree of Learning Upon The Relility of Nonsense Syllable Scores. Pub. Kansas State Teachers College, Emporia, 1932 p.8

to the wide distribution in the abilities of the subjects used in the present study. It is interesting to note that the correlation coefficient for the longer list is higher than that of the other two lists. This is not due to pratice because the counter-balance control was used. The P. E. 's are quite low for such high correlations. The true correlation lies between .738 and .871.

Table II shows the correlations between the comparable time scores for the three different lenghts of lists employed.

TABLE II

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

Lengths	"r"	P.E.
Six	.76	.0466
Twelve	.85	.0295
Eighteen	.835	.0331

Read table thus: The correlation in
time between the learning of the two
groups of six nonsense syllables is
.76±.0466

The P. E.'s in these correlation are higher than those found for trials. Nevertheless the coefficients possess a

high degree of reliability. Both the time and trial correlation are in close agreement; the range in the trial correlation is .133 and that of time is .156. There is not enough difference in the two to make a supposition as to which is the more reliable criterion of learning. In the case of the time scores the twelve-syllable lists are more closely correlated, by a small degree, than the eighteen-syllable list. Either appears to be more reliable than the six-syllable lists.

In order to test these correlations between comparable lengths of lists, the writer computed a correlation between the odd numbered syllables and the even numbered syllables in each length of list. Table III shows the results.

TABLE III

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.
Six	.86	.0201
Twelve	.96	.005
Eighteen	.97	.0047

Read table thus: The correlation Coefficient between the average number of correct recalls between the even trial and the odd trials for the six syllable is .86±.0201

In this case the list of eighteen gives the highest correlation; it is closely followed by the list of twelve and then six. The list of six has had the lowest correlation in two cases and the same as the list of twelve in the other. The list of twelve and eighteen in the trials to learn and odd against even trials have had very similar correlations. While the correlations of Table III do not represent true reliability coeffecients, they are valid for the purpose of showing the relationship between length and reliability.

The correlation coefficients, in the foregoing tables, are higher than those found by Lehman. In this experiment the group was more hetrogenous in abilities than Lehman's experimental group. It is a known fact that a wide distribution of abilities usually gives a higher correlation than a more restricted range. Furthermore Lehman used perhaps a too limited number class intervals in his correlations. His correlations were raised by using a greater number of intervals.

(2) THE EFFECT OF PRATICE ON THE RELIABILITY OF NONSENSE SYLLABLE SCORES

The relation of the much discussed problem of pratice in its relation to measurement in learning shown in Table IV. This table shows correlation of trials between first

day and second, third day and fourth, and fifth day and
sixth for all lengths of material.

TABLE IV

CORRELATION COEFFICIENT FOR TRIALS
ON SUCCESSIVE DAYS FOR THE ENTIRE
MATERIAL, REGARDLESS OF LENGTH

Days	"r"	P. E.
1st vs. 2nd.	.89	.0221
3rd vs. 4th.	.92	.0168
5th. vs.6th.	.93	.0154

Read table thus: The correlation co-
efficient between the entire material
for the first and second day is .89±.0221.

The correlations given in Table IV are abnormally
high owing to a spurious factor of variability in length
of material, since all lengths of material were used.
Nevertheless, this procedure is valid for our purpose
namely: showing the relation between pratice and reli-
ability.

In Table V the correlation coefficients for time,
between first and second days, third and fourth days,
and fifth an sixth days are given.

TABLE V

CORRELATION COEFFICIENTS FOR TIME ON
SUCESSIVE DAYS FOR THE ENTIRE
MATERIAL, REGARDLESS OF LENGTH

Days	"r"	P. E.
1st vs. 2nd	.88	.0241
3rd vs. 4th	.96	.0080
5th vs. 6th	.89	.0021

Read table thus: The correlation co-
efficient between the entire material
for the first and second days .88±.0241

Attention is called to the fact that the same spurious

factor exists in the case of the correlations between time

scores. In general these correlations fail to indicate

any greater reliability due to pratice.

In the following table (Table VI) the odd numbered

syllables are correlated with even numbered syllables on

different days.

TABLE VI

CORRELATION COEFFICIENTS SHOWING
CORRELATION OF ODD NUMBERED SYLLABLES
VERSUS EVEN NUMBERED SYLLABLES
ON CERTAIN DAYS

Syllables	Day	"r"	P. E.
Odd vs. Even	1st.	.97	.006
Odd vs. Even	3rd.	.97	.006
Odd vs. Even	6th.	.97	.004

Read table thus: The correlation coefficient of
odd numbered syllables versus even numbered syll-
ables on the first day was .97±.006

These days were selected arbitarily. It was more or
less a chance selection. The coefficients remain high
throughout. By selecting these days we have the reliabil-
ity on both odd and even days. The correlation indicate
that pratice has little effect on the reliability of non-
sense syllable scores.

Rhos were computed for each of the different lengths
of material, between first and second days, third and four-
th days, and fifth and sixth days for both time and trial.
They were found to be wide in their range. The results
are shown in Table VIII.

TABLE VII

RHOS OF DIFFERENT LENGTHS OF
MATERIAL FOR BOTH TIME AND TRIALS
ON SUCESSIVE DAYS

Length of List	Days	Time	Trial
Six	1st. vs. 2nd.	.66	.52
Six	3rd vs. 4th.	.63	.36
Six	5th. vs. 6th.	.36	.86
Twelve	1st. vs. 2nd.	.86	.91
Twelve	3rd. vs. 4th.	.68	.77
Twelve	5th. vs. 6th.	.99	.87
Eighteen	1st. vs. 2nd.	.81	.67
Eighteen	3th. vs. 4th.	.68	.87
Eighteen	5th. vs. 6th.	.78	.76

Read table thus; The rho for the list of six
between first and second day is .66 and for
trial is .52.

These correlations do not show that pratice aids in
the reliability of nonsense syllable scores. The rho for
time of the fifth and sixth day is lower than the one for
the first and second days. None of these rhos shows a
marked influence of pratice upon the reliability of non-
sense syllables. These correlations corrobarate the pro-
duct moment coefficients shown in Table V. Altogether the
rhos indicate fair reliability for the nonsense syllable.
These results confirm those of Lehman with respect to re-
lation between pratice and reliability.

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

The degree of learning was studied by correlating
the number of trials on comparable lists in which all but
two syllables were given correctly, then all but one syll-
able correct, then all right, then all right in sucession.
Another way of determining the effect of learning upon
reliability was by correlating day 1, 3 and 5 with days
2, 4 and 6 with all but two syllable correct, all but one
correct, all right, all right twice in sucession. Table
VIII shows the correlation coeffecients for the comparable
list of material.

TABLE VIII   (Next Page)

TABLE VII

CORRELATION COEFFICIENTS FOR COMPARABLE
LIST SHOWING DEGREE OF LEARNING
ON RELIABILITY OF NONSENSE
SYLLABLE SCORES

Length of List	Degree of Learning	"r"	P. E.
Six	All right but two	.42	.064
Six	All right but one	.64	.056
Six	All right	.74	.013
Six	All right twice	.71	.053
Twelve	All right but two	.68	.057
Twelve	All right but one	.74	.035
Twelve	All right	.74	.013
Twelve	All right twice	.75	.046
Eighteen	All right but two	.84	.031
Eighteen	All right but one	.83	.024
Eighteen	All right	.84	.031
Eighteen	All right twice	.86	.027

Read table thus: The correlation coefficient for degree
of learning on reliability of nonsense syllable scores
for list of six with all right but two is .421±.064

This data appear to show a slight tendency for reliability
to increase with degree of learning, especially in the case
of the shortest lists.

In these coefficients the highest one found consistently

between the lists of eighteen; the lists of six have the lowest. This would tend to show that the highest reliability in degree of learning is found in the longest lists and that reliability is roughly proportional to length of list.

Table IX shows the relation between reliability and degree of learning when the scores of days 1, 3 and 5 are correlated with those of days 2, 4 and 6.

TABLE IX

CORRELATION COEFFICIENTS OF THE
ODD DAYS VERSUS THE EVEN DAYS FOR
ENTIRE MATERIAL WITH DIFFERENT
DEGREES OF LEARNING

Days	Degree of learning	"r"	P. E.
1, 3 and 5 vs. 2, 4 and 6	All but two right	.83	.0194
1, 3 and 5 vs. 2, 4 and 6	All but one right	.91	.0107
1, 3 and 5 vs. 2, 4 and 6	All right	.89	.0127
1, 3 and 5 vs. 2, 4 and 6	All right twice	.91	.0107

Read table thus: The correlation coefficient of the odd days versus the even days with all right but two syllables .83±.0194

The coefficients shown in this table are consistently high owing to the spurious factor mentioned above. These results also for the greater degree of learning as indicated by the comparatively low coefficient for the scores all right but two.

(4)Composite Results of Lehman's and

McCue's Studies.

(a) The Effect of Length of List Upon The Reliability

of Nonsense Syllable Scores.

Table X

CORRELATION BETWEEN THE SCORES OF
PAIRED LISTS OF NONSENSE SYLLABLES
FOR EACH LENGTH OF LIST

Variables	Trials		Time	
	r	P. E.	r	P. E.
Six with Six	.61	.048	.68	.041
Twelve with Twelve	.73	.036	.87	.018
Eighteen with Eighteen	.77	.031	.62	.048

Read table thus: The correlation coefficient of the
paired list of six trials is .61±.048

With one notable exception these results point to a

greater reliability of the longer lists. It is significant

that in the case of both time and trial scores the coefficients

for the twelve-syllable list are marked by greater than

those of the six syllable lists. It is not clear as to

whether the reliability of the time scores diminishes at the

greater lengths, as r .62 would indicate, or whether

this low correlation is due to chance.

Since each subject learned two lists each of six, twelve and eighteen syllables, one hundred and fifty two cases are represented in each of the correlations between odd and even numbered items. TableXI gives the correlations for each of the three lengths used.

TABLE XI

CORRELATIONS BETWEEN NUMBER OF
TRIALS REQUIRED TO LEARN ODD AND
EVEN NUMBERED SYLLABLES

Variables	r	P. E.
Odd vs. even 6's	.85	.015
Odd vs. even 12's	.95	.005
Odd vs. even 18's	.96	.005

Read table thus: The correlation coefficient
for the odd numbered syllables and even num-
bered syllables for the list of six is .85±.015.

As previously mentioned the procedure give spurious correlations. They are not to be taken as true reliability coefficients.

As a final means of studying the relation between length of list reliability, P. E.'s and the ratios of the P. E.'s to their respective means were computed for the various lists of syllables used. The results for both time and trial scores are given in the following table.

TABLE XII

THE MEANS, P.E.'S AND RATIOS OF
P.E.'S TO MEANS FOR EACH LIST

Lists	Trials			Time in Minutes.		
	M	P. E.	M/Pe	M	P. E.	M/pe
1st. Six	6.00	.15	40.00	3.25	.16	20.31
2nd. Six	5.60	.16	35.00	2.80	.13	21.54
1st. Twelve	10.50	.28	37.50	18.24	.75	24.32
2nd. Twelve	10.11	.27	37.44	16.17	.68	23.78
1st. eighteen	14.92	.30	49.73	32.25	.78	40.06
2nd.Eighteen	1379	.33	41.78	26.91	.81	33.59

Read table thus: The mean for trial for the first list of
six is 6.00 with PE of .15. The ratio of the mean to P.E
is 40.00

These results show less variability, in proportion to
the size of the mean, for the longer lists in 11 out of 12
comparisons attention is called to the fact that the P.E.'s
of the eighteen syllable lists are very little larger than
those of the twelve-syllable lists. This,of course, makes
M/P.E much larger than that of the twelve syllable lists
and indicates greater reliability for the longer lists.
As will be mentioned later, the subjects were divided into
several smaller groups and separate means and σ's computed

for the scores of each group. In most cases the σ's of
the group.of eighteen-syllable lists were found to be
only slightly higher than those of the twelve syllable lists.

It is interesting to note that in every case but one
the ratio of P. E to M is larger for the first list of a
pair learned than for the second. This tends to show that
pratice of this amount increase group variability.

(b) The Effect of Pratice Upon the Reliability of Nonsense Syllable Scores

The first method of studying this problem consist of correlating the scores for the first day with those of the second day, the scores of the third day with those of the fourth and those of the fifth with those of the sixth. The amount of pratice represented in the first four days is as much as that usually required preliminary to participation in psychological experiments. Its influence ought to be apparent in the correlations between the scores of the fifth and sixth days, if it has any effect upon reliability. Each correlation given in the following table is based upon paired scores of the three lengths that is difference in lengths of lists are ignored. This, of course, is conducive to abnormally high correlations. However, since this spurious factor is constant it is not a limitation to the procedure herein employed. Seventy six cases are represented in each of the correlations.

TABLE XIII

CORRELATIONS BETWEEN SCORES ON
SUCESSIVE DAYS REGARDLESS
OF LENGTH

Variables	Trials		Time	
	r	P. E.	R	P. E.
1st Day vs. 2nd	.88	.017	.90	.014
3rd. Day vs. 4th	.86	.020	.86	.020
5th Day vs. 6th	.88	.017	.92	.012

Read table thus: the correlation coefficient for the first
day and second day for trial is .88$\pm$.017
day and second day

These data fail to show any effort upon reliability
from pratice.

In a further attempt to determine the effect of
pratice upon reliability the trials required to learn
odd and even numbered syllables were correlated for each
of the six days. Correlations of.96,.94,.97,.96,.97,
and .96 were obtained for days 1, 2 , 3, 4, 5, 6, respec-
tively. These coefficients fail to show any advantage
from pratice, so far as reliability of scores is concerned.

(c) The Effect of The Degree of Learning Upon the Reliability of Nonsense Syllable Scores.

The relation between the degree of learning and the reliability of the scores was studied by the correlational procedure. This was done by the correlating the trials required to learn comparable lists to the following degrees: (1) all right but two, (2) all right but one, (3) all right and (4) all right twice in successions. Table XIv shows the results.

TABLE XIV

THE RELATION BETWEEN DEGREE
OF LEARNING AND RELIABILITY
OF SCORES

Degree of Learning	Trials					
	r(6&6)	P.E.	r(12&12)	P.E.	r(18&18)	P.E.
Right but two	.42	.064	.74	.035	.81	.026
Right but one	.57	.051	.74	.035	.83	.024
All Right	.61	.048	.73	.036	.77	.031
All right twice	.64	.046	.75	.034	.80	.027

Read table thus: The correlation coefficient for the degree of learning of all right but two for the six syllable list is .42. .064

These results indicate that in the case of the six syllable list the greater the degree of learning required

the more reliable the scores are. The reliability co-efficients of the twelve and eighteen syllable list seems to be unaffected by the degree of learning. Coefficients were not computed for scores of lists learned to the point of three consecutive, errorless repetitions. There were only two subject who failed to achieve the third correct repetition after having achieved the second. Consequently, the correlations would be pratically the same as those under the conditions, "all right twice in succession", Table XIV.

These results confirm to those given above relative to the relation of length of list to reliability. All these correlations, regardless of degree of learning indicate greater reliability for the longer lists. The correlations also show considerable stability.

(d) Further Considerations of Reliability

Of Nonsense Syllable Scores

It seems that there can be no better measure of reliability than constancy of results. There are two questions involved . The first has to do with variability of the individual from one experiment to another. The second deals with variability from one group to another. For the purpose of study, the subjects were divided by chance selection into smaller groups of thirty eight, nineteen and six or seven as shown in Table XV. These results are in terms of trial scores. The object was two fold: (1) to see with what constancy the second performance of a group could be predicted from their first performance, and (2) to determine the variation between groups selected by chance.

TABLE XV

VARIATIONS IN MEANS AND P.E.'S
OF GROUPS OF SUBJECTS
SELECTED AT RANDOM

Subjects	(1)18	2nd18	1st12	2nd12	1st6	2nd6
1-38	15.2±.37	13.9±.42	10.6±.41	9.9±.36	5.8±.21	5.6±.19
39-76	14.5±.49	13.5±53	10.4±.39	10.3±.37	5.6±.19	5.6±.19
1-19	15.7±.54	15.3±.65	11.1±.67	10.3±.49	5.7±.27	5.8±.29
20-38	14.8±.48	13.1±.45	10.1±.46	9.6±.53	5.8±.32	5.3±.24
39-57	13.8±.59	12.7±.66	10.8±.62	9.5±.41	5.5±.21	5.1±.21
58-76	15.2±.21	14.2±.86	10.8±1.04	11.0±.58	6.8±.34	6.1±.45
1-6	14.7±.90	16.24	10.7±1.26	9.5±32	5.7±.44	6.4±.44
7-12	15.2±.75	14.2±1.25	12.1±1.10	9.8±87	5.5±.41	6.1±.34
13-18	17.0±.93	15.6±.96	9.7±.93	11.3±.58	6.0±.54	5.6±66
19-24	13.3±.79	12.7±.51	9.0±.67	9.3±1.03	4.8±37	4.7±.44
25-30	16.7±.67	11.8±.70	9.0±.67	8.7±.67	6.3±72	5.5±38
31-36	15.5±.89	14.4±.85	11.3±.72	10.7±.86	6.3±44	5.7±40
37-43	15.0±.87	15.0±1.31	11.0±.75	9.7±.64	5.5±31	4.8±41
44-50	13.23±.74	12.5±.80	10.2±.69	10.7±.70	5.8±30	6.3±.21
51-56	13.0±80	10.9±.79	9.1±.92	8.4±.65	5.1±60	4.1±.16
57-63	16.59±1.93	15.2±1.85	11.5±1.23	11.0±1.32	6.5±75	6.0±.27
64-70	13.0±35	12.5±.75	9.2±.55	9.3±.54	6.5±21	6.0±22
71-76	16.59.	14.8±1.71	11.6±1.12	12.4±1.2	7.3±61	6.3±42
Grand Ave.	14.9	13.8	10.5	10.1	5.9	5.6

Read table thus: the mean of subjects one to thirty eight for

first list of list has a mean of 15.2 with P.E. of .37

Inspection of table XV shows satisfactory correspondence between the performances upon comparable lists. The average difference between the scores on the first and second eighteen syllable lise is 1.1 trial, that between the two twelve syllable list is .4 of a trial, that between the six syllable is .3 of a trial. These differences may be due to pratice, since differences difficulty of lists of the same length controled . Other chance factors out cancel each other so that the constant improvement shown on the second list of each pair may be regarded as due to pratice. If one were to attempt to predict the second score of a pair from the score on the first pair there would be a constant error equal to the pratice effect. In the case of the eighteen syllable lists, for example, in order to compute the error due to pratice, one should subtract 1.1(pratice effort) from the first score and determine the difference between the product obtained and the second score. When the first score equals 15.5 and the second 14.4, 15.5-1.1-14.4 equal 0, the error of prediction. The following observations represent errors of prediction of a group of subjects from the previous performances of same group. The larger the group, if selected by chance, the smaller the error should be. With a sufficiently large group, if therebe no constant errors, there

ought to be no error in prediction. When the subjects were
divided into two groups of thirty eight each the average er-
ror in predicting the mean score on the second list from the
mean score on the first list when correction is made for
pratice is .15, .30 and .10 of a trial for the eighteen,
twelve and six list respectively. When the seventy-six sub-
jects are divided into four group prediction is .36, .32 and
.30 of a trial for the respective lists. Dividing the sub-
jects into twelve groups of six and seven the average error
in group prediction, corrected for pratice effect is .89
(median .32), .46 and .46 trials for the eighteen, twelve
and six syllable lists. In these computations the sign is
ignored. These data indicate fair predicability of perform-
ance of groups as small as six and seven subject, when their
previous performance is know. They likewise suggest that
nonsense syllables scores are fairly reliable.

Finally attention is called to the similarity between
scores of different group and the same lists. Inspection
of Table XV shows that most group averages correspond farily
close to the grand average for the 38 subject none varies from
the grand average by as much as three times their P.E. In

the 19 subject group, two of the twenty four means deviate

from the grand average by as much as three times their P.E.
There are seventy two small groups of six and seven subject
each. There are eight case in which the group means deviate
from their respective grand average by three or more times
their P. E. On a basis of a single random selection of the
small group these deviations are not be be regarded as
extravagant.

This completes the composite study of the two problems.
The larger number of cases confirm the results obtained in
each of the studies.

CONCLUSIONS

1. The reliability of nonsense syllable scores is sinificant-
ly influenced by the length of the lists.

    (a) All methods of correlations showed that the
.12 and 18 syllable lists are more reliabile than
the 6 syllable lists. The show with one except-
ion greater reliability for the 18 syllable
list than for the twelve list, altho the dif-
ferent is not marked.

    (b) The ratios of P.E.'s to means indicates with
single exception, consistent significant dif-
ferences in realiability in favor of the great-
er length. By this method the 18 syllable lists
appear to be significantly more reliable than
the 12 syllable list.

2. Pratice of the amount achieved in this study, so far as
these data are concerned, has no affect upon the reliabil-
ity of nonsense syllable scores.

3. These results show an improvement in reliablity of scores
of the shorter list with degree of learning up to the de-
gree of "all right once." Degree of learning does not af-
fect reliability of the eighteen syllable list. Learning

beyond "all right" does not affect reliability even in a shorter list.

4. The study as a whole suggest that nonsense syllables constitute a reasonable reliable instrument for experimental work.

BIBLIOGRAPHY

Ebbinghouse, Herman, Memory, A Contribution to Experimental
        Psychology, Translated by Ruger and Bussenius,
        Pub. Teachers College, Columbia University,
        N.Y., 1913, pp.81-120.

Lehman, Adin F. A Study of The Effect Of Length, Pratice
        And Degree of Learning Upon Reliability Of
        Nonsense Syllable Scores, Pub. Kansas State
        Teachers College, Emporia, July 1932, pp.1-37.

Leeper, Robert "The Reliability and Validity of Maze Experi-
        ments With White Rats", in GENETIC PSYCHOLOGY
        MONOGRAPHS, 1932, pp.137-245.

Robinson, E. S., and Heron ,W. T. "Results of Variation in
        Length of Memorized Material," JOURNAL OF PSYCH-
        OLOGY, V. 5, 1922, pp. 428-449.

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

Thurstone, L. L. "The Relation Between Learning Time and Length
        of Task", in PSYCHOLOGICAL REVIEW, 1930, 37,
        pp. 44-53

Tomilinson, Charles, Studies in Memory, Published Kansas State
        Teachers College, Emporia, 1932, pp. 1-47

APPENDIX

Materials. The following twelve lists compromise
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	List
(3)	(9)
GOX	BIR
BEF	FUB
XAB	CAH
NUR	KIR
DOK	NUV
FET	SEB

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

Group No.	1st Day Lists	2nd Day Lists	3rd Day Lists	4th Day Lists	5th. Day Lists	6th Day Lists
1	1 2 3	7 8 9	4 5	10 11	6	12
11	5 6	11 12	1	7	2 3 4	8 9 10
111	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