Image Cover Sheet

CLASSIFICATION	SYSTEM NUMBER 210144
UNCLASSIFIED	

TITLE

A STUDY OF RATE OF LEARNING AND CONSTISTENCY OF PERFORMANCE IN PERCEPTION OF SIMULATED RADAR BREAKS AND SIGNALLING THEM BY MEABS OF A MANUALLY CONTROLLED STY

System Number:

Patron Number:

Requester:

Notes:

DSIS Use only:

Deliver to:

COPY NO. 1

DEFENCE RESEARCH BOARD

DEPARTMENT OF NATIONAL DEFENCE

CANADA

A STUDY OF RATE OF LEARNING AND CONSISTENCY OF PERFORMANCE IN PERCEPTION OF SIMULATED RADAR BREAKS AND SIGNALLING THEM BY MEANS OF A MANUALLY CONTROLLED STYLUS

BY
C. H. BAKER
J. C. OGILVIE

DRML PROJECT NO. 61-34-4
DRML REPORT NO. 61

H.R. 39

15M—9-50 (56-507)

OTTAWA
10 SEPTEMBER, 1951

[redacted]

DEFENCE RESEARCH BOARD

DEPARTMENT OF NATIONAL DEFENCE
CANADA

A STUDY OF RATE OF LEARNING AND CONSISTENCY OF PERFORMANCE IN PERCEPTION OF SIMULATED RADAR BREAKS AND SIGNALLING THEM BY MEANS OF A MANUALLY CONTROLLED STYLUS

by

C.H. Baker and J.C. Ogilvie

DRML Project No. 61-34-4

DRML Report No. 61

OTTAWA

HR 39

May, 1952

TABLE OF CONTENTS

A STUDY OF RATE OF LEARNING AND CONSISTENCY OF PERFORMANCE IN PERCEPTION OF SIMULATED RADAR BREAKS AND SIGNALLING THEM BY MEANS OF A MANUALLY CONTROLLED STYLUS

A STUDY OF RATE OF LEARNING AND CONSISTENCY OF PERFORMANCE IN PERCEPTION OF SIMULATED RADAR BREAKS AND SIGNALLING THEM BY MEANS OF A MANUALLY CONTROLLED STYLUS

Abstract

This experiment was designed to determine whether learning was a factor in electrically signalling radar screen 'breaks' by contacting screen 'breaks' with a metal stylus under speeded conditions. The apparatus developed was considered, too, as a psychomotor test and an additional aim of the experiment was to determine the reliability of this test.

It was found that learning was extremely rapid on both of two practice days and that there was a carry-over of learning from the first to the second day, in both speed and accuracy. Considered as a short psychomotor test, scores made on the apparatus were found to be acceptably reliable.

INTRODUCTION

The usual method of reporting 'breaks' on a Plan Projection Indicator (PPI) tube is to have the radar screen teller telephone the positions of radar screen 'breaks' in terms of coordinates, with reference to a rectangular grid system superimposed upon the radar screen. This technique is slow; a maximum of six coordinates ('breaks') can be reported per minute by a trained operator. It is also inaccurate, being subject to occasional errors of coordinate inversion in reporting (and in receiving) and also to errors of central tendency when positions of 'breaks' between grid lines are interpolated (2).

Any system which makes possible faster and more accurate 'breaks' reporting warrants investigation. Such a system has been proposed. Basically, it consists of merely contacting screen 'breaks' with a metal stylus or strobe. Such contacts automatically signal the presence and accurate location of a 'break' to a central display.

At the time this study was undertaken it was anticipated that some studies of the effects of anoxia upon psychomotor performance might be undertaken in the near future. Could the preliminary study be designed in such a manner that the apparatus involved might provide a reliable test of psychomotor performance under conditions of anoxia? It appeared that it could.

Accordingly, the study reported here was designed to answer two questions:

(1) What learning is involved before an operator can employ such a 'break' reporting system efficiently?

(2) In using such a machine as a psychomotor test, how reliable are the scores on such a test?

APPARATUS

The apparatus was designed to simulate, in a general manner, the screen of a ten inch PPI tube and all experimental runs were performed in a room darkened to approximate the illumination in a radar control room.

A circular disk of metal, ten inches in diameter, was mounted on a wooden frame at an angle of fifty degrees with the horizontal. The disk was randomly punched with 100 holes, each 1/16" in diameter. Four 6-8 volt light bulbs were mounted immediately below the metal disk. Thin, green cardboard sheets were cut to the same size and shape as the disk. Varying numbers of 3/16" holes were punched in the cardboard sheets so that they were coincident with certain of the holes in the disk. Thus a cardboard face with 10 coincident holes would, when placed over the metal disk, occlude all but 10 of the disk holes. By using differently punched cardboard sheets any number of the disk holes could be revealed. Because of the difference in size between cardboard and disk holes, wherever a disk hole was revealed a small circle of the surrounding metal was also uncovered.

A short, threaded-end steel shaft was passed through the centre of the disk and cardboard face to a 4 rpm telechron motor mounted inside the wooden frame. A centrally punched black cardboard disk with a ninety degree sector removed was mounted on the disk end of the shaft by a wing nut. When the telechron motor operated, the sectored disk revolved at 4 rpm immediately in front of the metal disk (and cardboard face). The leading edge of the sectored portion of the revolving disk continuously revealed 'new' illuminated holes (which simulated screen 'breaks') while the trailing edge subsequently occluded them.

A wooden six inch metal-tipped stylus with a metal core was wired into a 12 volt circuit. At the forefinger position of the stylus (when held as a pencil) a small switch was installed. Two Cenco counters were wired into the same circuit. When the stylus tip contacted the disk hole (but did not penetrate, being of slightly larger diameter), and the switch pressed, both counters were activated. When, however, the stylus contacted the cardboard face rather than a disk hole or the metal area around it only one counter was activated (Fig. 1). Thus, stylus contact with a disk hole or its revealed metal circumference was considered a correct contact (both counters activated). When the stylus was inaccurately aimed at a disk hole and missed, contacting instead the cardboard face, only one of the counters was activated. The difference between readings of the counters over a period of time, then, could be considered an error score.

A possible source of error in the counting system was that subjects could hold the switch down continuously while making contacts. To check such errors a small red bulb was wired into the counter circuit. The bulb was illuminated and extinguished at every switch press and release. If the bulb was steadily illuminated it indicated to the Experimenter that the subject was not following instructions.

This apparatus makes possible an unusual number of experimental variations, some of which are below:

(1) Number of 'breaks' presented, which is determined by the number of coincident holes in the cardboard face.

(2) Length of life of 'breaks', which is determined by the angular size of the sector.

(3) Speed of sweep of the sectored disk.

(4) Area of the 'breaks', i.e. area within which stylus contacts can be designated as correct, which is determined by the diameter of the coincident holes in the cardboard face.

(5) Studies of the effects of prolonged visual searching and stylus manipulation, i.e. determination of work decrement in time.

(6) Studies of the effects of differently illuminated surrounds.

(7) Studies of operator accuracy when contacts are paced by a metronome.

(8) Studies of operator accuracy when working at top speed.

By wiring counters to any predetermined number of adjacent segments of the metal disk which are insulated from one another further variations are possible, e.g.

(9) Studies of the order of preference of central, medial, and peripheral contacts made by the operator when he has a choice.

(10) Studies of the order of preference of various sectors of the metal plate.

METHOD

General

When this study began there was no way of knowing whether learning would be apparent on the first few trials or a gradually accumulating phenomenon, i.e. learning might be revealed by increments of successful performance on successive trials in any one experimental session or by increments on successive days. Accordingly, subjects were run several times on each of two days.

Concerning the possible employment of the apparatus as a psychomotor test it was obvious that the test must discriminate between subjects. This consideration precluded the possibility of determining the top speed at which subjects could work with complete accuracy for two reasons: (1) it was felt that the range of speed scores would be small, and (2) there would be no range of Accuracy scores as all subjects would achieve perfect accuracy. One method of pointing up individual differences was to permit subjects to operate as they could and so yield both Speed and Accuracy scores. The test, then, would be a highly speeded one.

Specific

Subjects were brought into the laboratory where the operation and purpose of the apparatus was explained to them. The Experimenter then had the subject sit on a stool facing the apparatus and gave the subject the stylus in his preferred hand. The small illuminating lights were turned on and the sectored disk started revolving at 4 rpm. The subject was allowed a few trial runs in order to ensure that the task was thoroughly understood. For these runs 35 'breaks' were exposed, this number having previously been determined by the Experimenter as sufficient to keep subjects working at top speed with a sweep of 4 rpm and the sector size involved.

The Experimenter instructed the subject as follows: 'You are to go as fast as you can and be as accurate as you can'. The Experimenter sat in front of the counters in order to record scores during rest periods, dimmed the room lights, and gave the subject the 'Go' signal. The trials were timed with a stop watch. Each subject was given six two-minute trials with a two-minute rest period between trials. All subjects repeated the performance on a second day.

SUBJECTS

A disappointing aspect of this experiment was that very few subjects were available. Only twenty subjects (nineteen men and one woman) were obtained, all of them being civilian laboratory personnel between the ages of sixteen and thirty-two. Such subject poverty weakened the conclusions which could be drawn from the statistical analysis. Had more subjects been available the experimental plan would have been broadened considerably, but, under the circumstances it did not seem warranted.

RESULTS

A. Learning

It was decided that little was to be gained by reproducing the raw scores or the performance curves for every subject. In the interest of clarity, however, the raw data for one subject are reproduced below.

TABLE 1

Raw data for subject S.M.

	DAY 1			DAY 2		
Trial	Speed score. (Contacts)	Accuracy score. (Contacts)	D. Error score.	Speed score. (Contacts)	Accuracy score. (Contacts)	D. Error score.
1	189	174	15	184	182	2
2	178	177	1	225	210	15
3	198	189	9	225	208	17
4	210	199	11	228	210	18
5	209	195	14	242	209	33
6	211	200	11	257	212	45
Mean:	199.16	189.00	10.16	226.83	205.16	21.66

An analysis of variance was carried out on the Speed, Accuracy, and Error scores. In each case the analysis was done in steps, first for successive pairs of trials, then for each day, and finally for the two days together.

Where possible, the residuals from each pair of trials were pooled after testing for homogeneity of variance by Bartlett's method as given by Edwards (3). This was done in order to test the significance of interactions of subjects with, and differences between, pairs on each day, and the difference between the two days. The analysis of variance of Speed scores is shown in Table 2.

TABLE 2

Analysis of variance of Speed scores

Trials (Day 1)	1	2	3	4	5	6	Day 1.
Mean:	178.1	195.5	210.8	220.3	231.7	232.2	211.4
Trials (Day 2)	7	8	9	10	11	12	Day 2.
Mean:	198.0	220.0	230.5	244.6	251.1	260.8	234.2

Source of variation	df	Sum of Squares	Mean Square	F	Level of Significance
Between Trials: 1 & 2 (Pair 1):	1	3045.0	3045.0	13.2	1%
3 & 4 (Pair 2):	1	902.5	902.5	3.9	Not sig.
5 & 6 (Pair 3):	1	2.5	2.5	---	
7 & 8 (Pair 4):	1	4950.6	4950.6	10.7	1%
9 & 10 (Pair 5):	1	1974.0	1974.0	4.3	5%
11 & 12 (Pair 6):	1	940.9	940.9	1.0	Not sig.
Between pairs 1, 2, & 3 (Day 1)	2	41541.4	20770.7	89.9	1%
Between pairs 4, 5, & 6 (Day 2)	2	44615.4	22307.4	48.1	1%
Between Days 1 and 2:	1	31076.5	31076.5	---	
Between Subjects (N=20):	19	566082.1	29793.7	---	
Interaction: Subjects with Day 1 pairs of trials:	38	34599.3	910.5	3.9	5%
Interaction: Subjects with Day 1 pairs of trials:	38	81752.9	2151.4	4.7	5%
Interaction: Subjects with Days:	19	153377.7	8972.5	No test	
# Pooled residuals for pairs of trials:					
Day 1:	57	13173.5	231.1		
Day 2:	57	26415.5	463.4		
TOTAL:	239	1004449.8			

Residuals for Day 1 are not significantly different. For Day 2 they are just significant at the 1% level but may reasonably be pooled. They cannot be pooled for both days, however, as they are significantly different at the .1% level.

The Speed scores show that there is a rather sharp initial rise which smooths off to a plateau. On Day 2 the initial rise is faster and does not smooth off as soon as it does on Day 1.

The residuals could be pooled for each day but not for both days combined. The F values shown were obtained by using the appropriate denominator. The interactions are significant at the 5% level, in that the differences between pairs of trials on each day and between the two days are dependent upon the subjects, i.e. given a different group of subjects the results might be different. However, these interactions are not high, being significant at the 5% level only.

The interaction for the two days cannot be tested, nor can the mean square for the difference between the two days, but by inspection it is obviously significant.

The analysis of variance of Accuracy scores is shown in Table 3.

TABLE 3

Analysis of variance of Accuracy scores

Trials	1	2	3	4	5	6	Day 1
Mean	154.1	168.4	177.9	183.6	189.2	189.6	177.1
Trials	7	8	9	10	11	12	Day 2
Mean	167.7	183.6	190.2	192.3	192.4	192.6	186.5

Source of variation	df	Sum of squares	Mean square	F
Between trials 1 & 2 (Pair 1)	1	2044.9	2044.9	17.9 1%
3 & 4 " 2	1	424.9	424.9	3.7 ns
5 & 6 " 3	1	1.3	1.3	----
7 & 8 " 4	1	2512.3	2512.3	22.0 1%
9 & 10 " 5	1	42.1	42.1	----
11 & 12 " 6	1	0.3	0.3	----
Between pairs 1, 2 & 3 of day 1	2	16557.8	8278.9	72.5 1%
" " 4, 5 & 6 of day 2	2	7051.1	3525.6	30.9 1%
Between day 1 and day 2	1	5245.2	5245.2	46.0 1%
Between subjects (N=20)	19	220444.6	11602.3	
Interaction: Ss with pairs of trials: day 1	38	7591.3	199.8	1.8 1%
day 2	38	14663.2	385.9	3.4 1%
Interaction of Ss with days	19	27889.7	1467.9	12.9 1%
Pooled residuals #	114	13002.2	114.1	
TOTAL	239	317470.9		

The residuals are just significantly different at the 5% level but may reasonably be pooled.

With reference to the Accuracy data (Table 3) it can be seen that there is a rapid initial rise in Accuracy on each day to a plateau, The significant improvement on Day 2 is due chiefly to the higher starting level. The significance of the interactions shows that the differences between pairs of trials on each day, and between the two days, are again dependent upon the subjects. These interactions are barely significant at the 1% level.

It should be noted that the greater part of the total variation is due to the difference between subjects, indicating that the test is a reliable one for distinguishing between subjects when Accuracy scores are used.

The Error scores are not independent of the Speed and Accuracy scores but an analysis reveals some interesting points. The analysis of variance of Error scores is shown in Table 4.

TABLE 4

Analysis of variance of Error scores

ERRORS								
	Trials	1	2	3	4	5	6	Day 1
	Mean	24.4	27.3	32.9	36.7	42.5	42.6	34.4
	Trials	7	8	9	10	11	12	Day 2
	Mean	30.2	36.7	40.3	52.3	58.7	68.3	47.7

Significance of F for individual analyses

Sources of variation		df	F
Between 12 trials (days 1 & 2)		11	1%
" 6 "	day 1	5	1%
" 6 "	day 2	5	1%
Between trials	1 & 2	1	ns
	3 & 4	1	ns
	5 & 6	1	ns
	7 & 8	1	5%
	9 & 10	1	5%
	11 & 12	1	ns

[1] Since the residuals are very significantly different no test can be made of interactions and so the full table of the analysis is not shown.

In Table 4 the residuals could not be pooled and so the results are shown in brief only. Although the differences between Trials on both days and on each day considered separately are significant at the 1% level, none of the differences between pairs is significant at better than the 5% level. The significances of the interactions in the foregoing analyses imply that there are large individual differences in subjects' responses to a single experimental situation, and also

great variation in the response of any one subject from day to day. Due to this interaction the residuals cannot be pooled. Successive trials do not show a significant increase in Error scores, although taken as a group the scores are not homogeneous, indicating a gradual though significant increase in Error scores over the twelve trials.

B. Relation of Speed to Accuracy

Coefficients of correlation were computed to determine the degree of relation between Speed and Accuracy for both experimental days. The scores employed in computing each coefficient were the mean scores of each of the six trials for each subject. Because of the small N available for product-moment computation both product-moment and rank-order coefficients were computed in the hope that they would be of similar magnitude and thus warrant more confidence being placed in them than in either one alone. The coefficients are shown below in Table 5.

TABLE 5

Showing correlation coefficients of the degree of relation between Speed and Accuracy

	Product-moment (r).	Rank-order (rho).
Day 1.	.78	.73
Day 2.	.78	.78

It is evident from the coefficients shown in Table 5 that there is a marked relationship between Speed and Accuracy scores for both days.

C. Reliability

It is recognized that any reliability coefficients derived from such a few number of cases might be seriously questioned. Consequently reliability was determined in several ways in the hope that any agreement among coefficients so obtained might serve to establish more faith in the several similar coefficients than would any one alone. It is recognized, too, that the split-half technique is in poor repute for determining the reliability of a highly speeded test. This criticism is dealt with in a later discussion.

Finally, it can be noted that subject poverty in this experiment did not seem to warrant the development of some technique for combining Speed and Accuracy scores into a single score for each individual. Instead, reliability coefficients were determined for both Speed and Accuracy scores.

1. Split-Half Reliability

The splitting, for both Speed and Accuracy scores, and for both days, was accomplished in two ways, viz. mean odd vs. mean even scores, and mean of trials 1, 2 and 3, vs. mean of trials 4, 5 and 6. In each case both rank-order and product-moment coefficients were computed. Product-moment coefficients were computed, too, between scores on trial 1 and scores on trial 2 for both days. The computed coefficients are shown in Table 6.

TABLE 6

Reliability coefficients computed using the split-half technique

	Splitting technique	rho		r		r corrected by Spearman-Brown formula.	
		Day 1	Day 2	Day 1	Day 2	Day 1	Day 2
SPEED	Mean of 3 odd trials vs. mean of 3 even trials:	.93	.99	.98	.97	.99	.98
SPEED	Mean of first 3 trials vs. mean of second 3 trials:	.59	.94	.73	.84	.84	.91
ACCURACY	Mean of 3 odd trials vs. mean of 3 even trials:	.92	.97	.94	.96	.97	.98
ACCURACY	Mean of first 3 trials vs. mean of second 3 trials:	.88	.86	.93	.85	.96	.91
SPEED	Score on trial 1 vs. score on trial 2:			.84	.95		
ACCURACY	Score on trial 1 vs. score on trial 2:			.86	.95		

It can be seen from Table 6 that the rank-order and product-moment coefficients are, generally speaking, of approximately similar magnitudes. With two exceptions, (rho = .59, and r = .73, both on Day 1) the correlation can be described as 'very high'.

2. *Test-Retest Reliability*

One argument against the use of test-retest reliability coefficients for speeded tests is that 'learning and memory are a source of variance, both true and error, in unknown proportions' (4). Another argument, which applies to this experiment specifically, is that the intervals between Day 1 and Day 2 (test-retest) were unavoidably of different lengths (from 1 to 23 days - mean interval of 8 days). Nevertheless, coefficients were computed for Speed and for Accuracy scores using mean trial scores as test scores. The results are shown in Table 7.

TABLE 7

Showing reliability coefficients computed using the test-retest technique

DAY	rho	r
Day 1	.82	.66
Day 2	.75	.77

As can be seen from an inspection of Table 7 the test-retest coefficients are considerably lower than those computed using the split-half technique (Table 6). The Table 7 relationship can at best be described as 'Marked'.

DISCUSSION

A. Learning

The group performance curves for Speed, Accuracy, and Error scores are revealed by an inspection of the mean trial scores at the tops of Tables 2, 3 and 4. With one exception (mean Error scores for Trial 2) the scores on each day grow progressively larger. It can be noted, too, that the beginning performance on Day 2 (Trials 7) yield approximately the same scores as the Trials 2 on Day 1.

The Speed score analysis indicates that Speed increases rapidly for the first few trials on Day 1 and then levels off during the later trials. On Day 2 subjects begin at a faster pace than on Day 1 and do not level off as quickly. Learning has taken place in the trials on each day and there is a carry-over from Day 1 to Day 2.

The Accuracy score analysis indicates a similar learning, i.e., rapid learning within the trials on each day with a smoothing off near the later trials, and a carry-over from Day 1 to Day 2.

The Error scores, not being independent of either the Speed or Accuracy scores, to some extent reflects the characteristics of the latter. It is evident from the Error score analysis (Table 4) that on Day 1 Speed and Accuracy scores accelerate at similar rates, while on Day 2 the Speed and Accuracy acceleration differs for the first four trials but approaches the same acceleration on the last two trials. It is evident, too, that Errors significantly increased each day.

It can be generally stated then, that for both Speed and Accuracy on Day 1 there is a rapid initial learning which approaches a plateau after three or more two-minute trials. On Day 2 subjects' initial performance is slightly superior to the initial performance on Day 1, due to some carry-over of learning from the first day to the second. On the first day Speed and Accuracy accelerate at the same rate, while on the second day they begin at different rates but by the end of four trials are once again accelerating at the same rates.

B. Reliability

This test can be best described as a highly speeded psychomotor test. It falls in the category of what Guilford calls a 'perceptual-speed test' having homogeneous items and hence measuring one ability (4).

It is axiomatic that the split-half technique is a poor technique for determining the reliability of speeded tests (1, 4, 5). However, a modification of the usual technique is considered acceptable. Adkins states, 'If performance per unit of time rather than the test item is treated as the unit, however, there are practical methods of applying the split-half approach to speed tests'. Guilford points out that, 'A good device to use in the development of new tests is to prepare two equivalent halves and to administer them in immediate succession as two separately timed tests'. Indeed, this is the technique which has been followed in this experiment. In splitting odd vs. even, or first half vs. second half, the two halves yield scores which were made over two similar periods of time in immediate succession. Such being the case then, it is considered that the large coefficients presented in Table 6 are reasonably accurate indices of the reliability of this test in the sense that, given one set of scores the others can be closely predicted.

The problem arises as to how scores should be predicted in a practical situation, viz. in an anoxia experiment the deviation of actual scores from predicted scores after subjects have been deprived of oxygen would be the result, presumably, of the oxygen lack. Probably the simplest method would be to take advantage of the Speed and Accuracy coefficients derived from correlating scores on trial 1 with scores on trial 2 (r, in both cases, for Day 2 is .95 - Table 6). Subjects would each be exposed to a Day 1 session. On anoxia day subjects would be given one two minute trial, followed by deprivation of oxygen, followed by (or simultaneously with) another two minute trial. This technique could be used with or without the addition of a control group.

Table 7 gives some coefficients which are not encouraging if considered as reliability coefficients. The arguments against the use of these coefficients have been presented on Page 10. They are interpreted to mean that when neither learning, memory, or time interval have been controlled there still exists a marked relationship between Day 1 and Day 2 mean scores.

CONCLUSIONS

1. In using a system which reports radar screen 'breaks' by means of screen contact with a manually controlled stylus subjects reach their peak Speed and Accuracy performance in a very few minutes. There is some accumulative learning from the first to the second day of performing this task.

2. The highly speeded psychomotor test described is acceptably reliable.

3. Working at top speed in the test described there is a marked relationship between Speed and Accuracy scores (r = .78).

REFERENCES

1. ADKINS, D.C., Construction and Analysis of Achievement Tests, U.S. Gov't Printing Office, Washington, D.C., 1947, P. 293.

2. BAKER, C.H., Distributions of Plot Coordinates reported by Radar Screen Tellers and Relayed by Movement Tellers, Defence Res. Med. Lab., Project No. 64-34-4 (1951) 1-8.

3. EDWARDS, A.L. Experimental Design in Psychological Research, N.Y., Rinehart, 1950, P. 195.

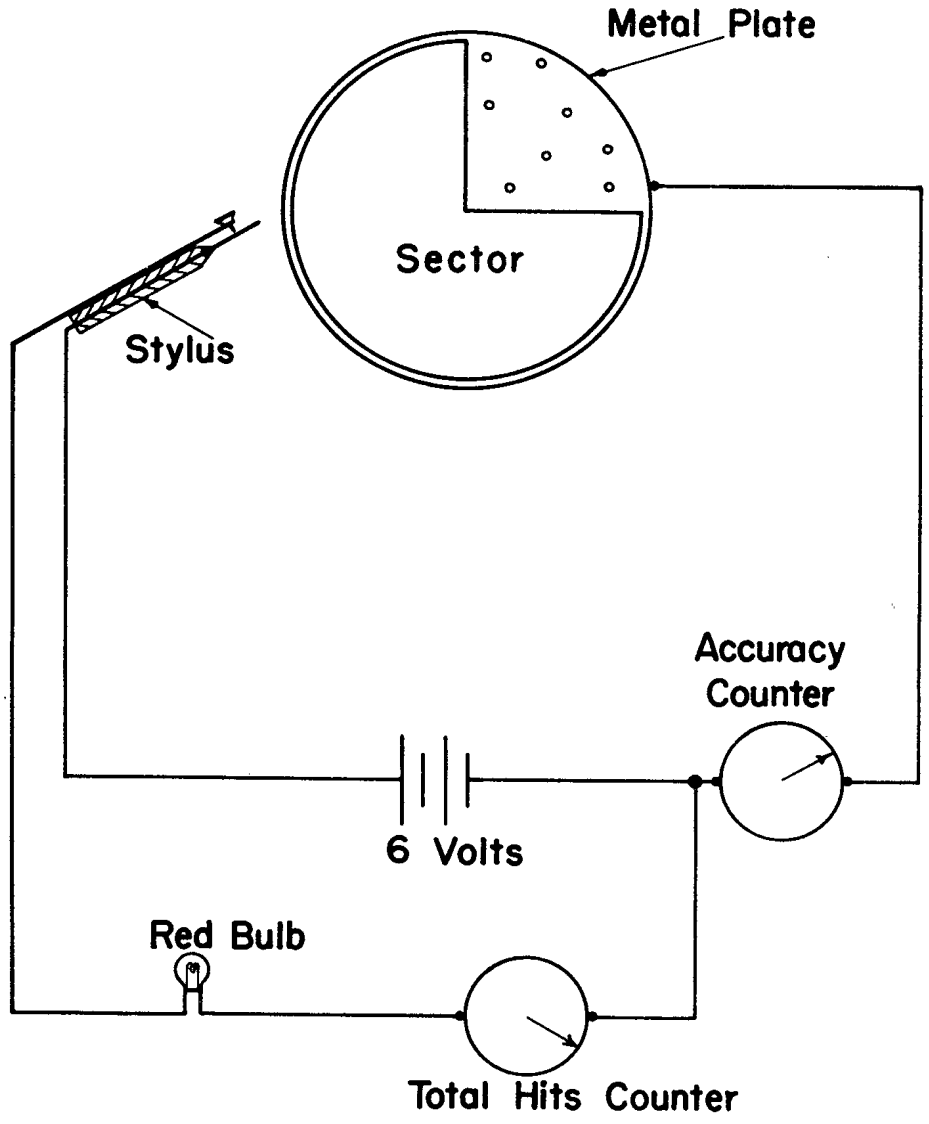
4. GUILFORD, J.P. Fundamental Statistics in Psychology and Education (2nd Ed.) N.Y., McGraw Hill, 1950, P. 633.

5. GULLIKSEN, H. Theory of Mental Tests, N.Y., Wiley, 1950, P. 486.

[redacted]

FIGURE 1

WIRING DIAGRAM



[redacted]

UNCLASSIFIED,
UNLIMITED

#210144