REPORT No. 5914, Part 4 February 1950

DESIGN OF INSTRUMENT DIALS FOR MAXIMUM LEGIBILITY

Part 4. Dial Graduation, Scale Range and Dial Size as Factors Affecting the Speed and Accuracy of Scale Reading

William E. Kappauf
William M. Smith

Princeton University

[redacted]

United States Air Force
Air Materiel Command
Wright-Patterson Air Force Base, Dayton, Ohio

NOTE

When drawings, specifications, and other data prepared by the War Department are furnished to manufacturers and others for use in the manufacture or purchase of supplies, or for any other purpose, the Government assumes no responsibility nor obligation whatever; and the furnishing of said data by the War Department is not to be regarded by implication or otherwise, or in any manner licensing the holder, or conveying any rights or permission to manufacture, use, or sell any patented inventions that may in any way be related thereto.

The information furnished herewith is made available for study upon the understanding that the Government's proprietary interests in and relating thereto shall not be impaired. It is desired that the Patent & Royalties Section, Office of the Judge Advocate, Air Materiel Command, Wright-Patterson AFB, Dayton, Ohio, be promptly notified of any apparent conflict between the Government's proprietary interests and those of others.

Espionage Act

Notice: "This document contains information affecting the national defense of the United States within the meaning of the Espionage Act, 50 U. S. C., 31 and 32, as amended. Its transmission or the revelation of its contents in any manner to an unauthorized person is prohibited by law."

"The above Espionage Notice can be disregarded unless this document is plainly marked with a security classification as "Restricted," "Confidential," "Secret," or "Top Secret."

The U. S. Government is absolved from any litigation which may ensue from the contractor's infringing on the foreign patent rights which may be involved.

AF TECHNICAL REPORT NO. 5914, Part 4 February 1950

DESIGN OF INSTRUMENT DIALS FOR MAXIMUM LEGIBILITY

Part 4. Dial Graduation, Scale Range and Dial Size as Factors Affecting the Speed and Accuracy of Scale Reading

William E. Kappauf
William M. Smith

Princeton University

United States Air Force
Air Materiel Command
Wright-Patterson Air Force Base, Dayton, Ohio

Air Force-WPAFB-(A)-O-23 JUN 50 620

FOREWORD

This report was prepared at Princeton University under USAF Contract No. W33-038 ac-14480 covering work on legibility problems in instrument design. The contract was initiated under the research and development project identified by Expenditure Order No. 694-15, and was administered by the Psychology Branch of the Aero Medical Laboratory, Engineering Division, Air Materiel Command, with Dr. Walter F. Grether acting as Project Engineer.

Three previous reports in this series from Princeton University have been published as USAF Air Materiel Command Memorandum Reports TSEAA-694-1L, 20 October 1947, MCREXD-694-1N, 12 July 1948, and USAF Technical Report No. 5914, (Part III), January 1950.

ABSTRACT

This study was conducted for the purpose of gathering quantitative evidence on the extent to which dial reading errors and reading times are influenced by the spacing of graduation marks, graduation mark values, scale range, and dial size.

The dials employed were scaled from 0 to 50, 0 to 100, 0 to 200, 0 to 400, and 0 to 600 units. Graduation schemes included graduations by tens, fives, twos, and units. Dial sizes were 2.8 inches and 1.4 inches. Twenty subjects read these dial materials on cards of 12 dials each under instructions to read to the nearest unit and to work as rapidly as possible. In all, 34,400 readings were involved.

The results show that the precision of scale reading, as measured by the frequency of errors of interpolation, rounding and the like, varies with the distance on the scale allocated to each scale unit. Reading precision improves as this distance increases to about 0.05 inches, a value which varies only slightly with graduation scheme but which appears to vary with graduation mark thickness. For more expanded scales, reading precision remains reasonably uniform at a level which depends on graduation scheme. Here, graduation by units or twos is better than graduation by fives, which in turn is better than graduation by tens.

Systematic or large scale reading errors seem to be determined by the same factors or conditions which determine reading time, namely, scale range and those aspects of graduation mark organization which must vary when longer and longer scale ranges are accommodated on a dial of a given size.

Although this abstract tends to focus attention on the direction of observed differences in error rates and reading times between dials of different design, the primary purpose of the report itself is to show the approximate magnitude of these differences. Such measured differences represent an essential part of the information needed by an engineer when weighing the advantages and disadvantages of alternative scale designs for particular instruments.

PUBLICATION APPROVAL

For the Commanding General:

Walter A. Carlson

WALTER A. CARLSON
Colonel, USAF (MC)
Chief, Aero Medical Laboratory
Engineering Division

CONTENTS

DESIGN OF INSTRUMENT DIALS FOR MAXIMUM LEGIBILITY.
IV. DIAL GRADUATION, SCALE RANGE AND DIAL SIZE
AS FACTORS AFFECTING THE SPEED AND ACCURACY
OF SCALE READING

I. INTRODUCTION

It is the responsibility of instrument designers to make practical decisions regarding the size of dial and kind of scale they will use in displaying the response of instruments which are to be read within certain stated tolerance limits. The scaling problem on any particular occasion may take one of several forms: If size of dial is fixed by other considerations, what is the greatest amount of quantitative information which can be crowded onto it? Or, if scale range is fixed by the nature of the information needed by the operator, how large must the instrument display be and how must the scale be graduated in order that readings can be made with the required precision?

The present experiment was set up for the purpose of obtaining data which would assist in answering questions of this sort. A series of dials was prepared covering likely scale ranges, typical graduation and marking schemes, and two commonly employed dial sizes. These were presented to subjects with average or better than average vision who were instructed to read them as rapidly as possible. The obtained records of reading speed and accuracy extend earlier data on the performance of subjects reading under accuracy instructions (Kappauf and Smith, 1948), and supplement the now rapidly growing literature on the effect of instrument design features on reading performance (Grether and Williams, 1947; Chapanis, 1947; Ford, 1949; Leyzorek, 1949).

When referred to in other reports of this series, the present experiment will be designated as Experiment 4. Nomenclature used in the present report includes the following: "scale range," referring to the number of scale units represented on the dial; "size of marked interval," referring to the spacing of graduation marks; and "size of called interval," referring to the scale distance representing the unit or units to which readings are made.

II. APPARATUS AND GENERAL PROCEDURE

The apparatus and testing procedure employed in this experiment have been described previously (Kappauf, Smith, and Bray, 1947; Kappauf and Smith, 1948). The subject sits within a dark test enclosure. With his head positioned at a 28 inch reading distance by a forehead rest, he reads serially a bank of 12 identically designed dials presented on a stimulus card. The central 10 of these are test dials. The experimenter records the called values on a tally sheet and takes stop clock time on the reading of the block of test dials. Prior to the illumination of each stimulus card, the subject views and is asked to describe a sample dial which indicates the type of scale graduation employed on the card about to be read. About 50 cards can be read conveniently in an hour of testing.

III. SUBJECTS

The subjects were 20 high school youths, ranging in age from 15 to 18 years. These 20 were screened from a larger available group on the basis of a visual acuity test administered in the dial reading situation under the same illumination provided for the dial cards. The Lebensohn near vision acuity chart (Lebensohn, 1936) was used. Inasmuch as the test was given at 28 inches instead of at the clinical near reading distance, the subject was required to read correctly eight of the 11 figures marked 20/40 on the chart in order to qualify as having at least 20/20 vision.

All subjects had been pretested on the Air Force Normative Test Battery, a series of pencil and paper tests covering arithmetic ability, graph reading, table reading, meter reading, distance estimation, vocabulary, etc. Scores on the various tests of this battery, as well as scores on the Henmon Nelson test of intelligence which is regularly administered at the high school from which the subjects were recruited, are listed in Table A of the Appendix.

IV. STIMULUS MATERIAL

The dials presented to the subjects were high contrast photographic prints on mat paper. As illuminated in the test situation, the white dial markings had a brightness of three foot lamberts. The black background was at about one-tenth this brightness.

Fifteen different dial designs were employed. These are shown in Figure 1. Choice of scale ranges between 50 and 600 units followed two original decisions: to use full 360 degree dials, and to restrict the study to scales which the subjects might reasonably be required to read to the nearest unit. Graduation schemes were the familiar ones involving markings every unit, every two units, five units, or ten units. Mark crowding limited the use of the two finest graduation schemes to dials with the shorter scale ranges.

The graduation mark pattern employed in going from major to intermediate to fine markings resembled patterns in use on current aircraft instruments. Because it was impractical to consider using a mark of the same specific length and thickness to designate a given numerical interval on all scales (as, for example, a standard 10 units mark for all dials), specific marks were allowed to vary from scale to scale as demanded by what appeared to be the most adequate mark pattern in each case. To the extent, then, that graduation mark thickness influences reading errors, the data to be discussed below will not always be an exclusive function of the separation or spacing of the graduation marks.

Choice of scale numbering was straightforward for the 50's, 100's, and 600's dials, but on the 200's and 400's dials various numbering arrangements would have been possible. Putting 10 numbers on each of the latter made them agree with the 100's dials with respect to number of scale numbers.

Dials with 2.8 inch scale diameter and 1.4 inch scale diameter were obtained by photo-enlargement from a common set of negatives. Dials of the two sizes therefore

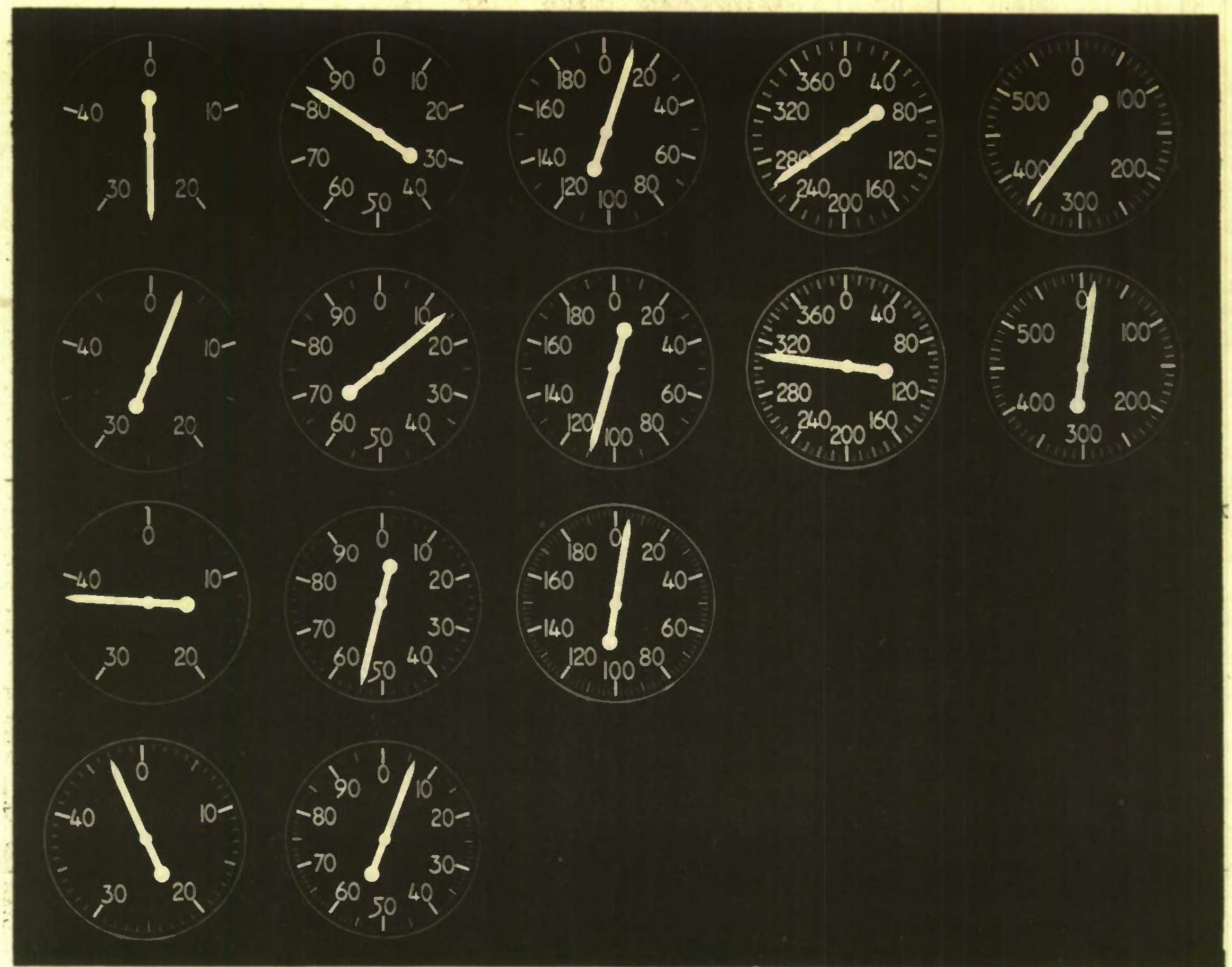

Figure 1. Types of dials used in the experiment. Note: These dials are "negatives" of those actually read in the experiment. They are shown here as black on white instead of white on black in order to make reproduction easier.

Descriptive data for 2.8 inch dials

All measurements are given in inches. Divide by 2 for 1.4 inch dials.

Graduation Marks:	Major	Intermed.	Fine	Numerals		Pointer	
Length	.22	.14	.08	Height	.27	Length from Center	1.32
Stroke Thickness	.045	.022	.011	St.Th.	.033	Width	.12

Length of arc devoted to one scale unit on each type of 2.8 inch dial

Dial	600	400	200	100	50
Arc length/unit	.014	.022	.044	.088	.176

differed in all scale dimensions -- graduation mark thickness, graduation mark length, pointer size, etc.*

For every type and size of 50's dials, every subject made 50 readings. Each unit scale position was used once on each dial. On each of the dials of other types, every subject made 60 readings. These were split systematically as follows: 30 on the right half of the dial and 30 on the left; six having 0 as the final digit, six having 1 as the final digit, etc.; six having 0 as the tens digit, six having 1 as the tens digit, etc.; and equal numbers of readings occurring within each 100 units of scale range. Within these restrictions, the particular number combinations used were obtained by working from random number tables. In the case of the 400's and 600's dials, once the pointer settings had been established for the dials graduated by five unit intervals, settings differing by exactly 180 degrees were used for the dials graduated by tens.

It should be noted that the pointer settings shown on all the dials were held within 0.3 units of some exact unit position on the scale (see Kappauf, Smith, and Bray, 1947). This control of setting was introduced in order to eliminate ambiguous readings at or near 0.5 unit positions and made it possible to be rigorous in classifying a misreading of one unit as an error. At the same time, of course, it operated to reduce somewhat the number of errors of interpolation which were made.

V. PLAN OF THE EXPERIMENT

Each subject came to the laboratory for six reading sessions. On the first day he was given the acuity test, familiarized with the test routine, and read a deck of practice cards covering all dial types to be used in the experiment. On the subsequent five days he spent one test session each on a 50's deck, a 100's deck, a 200's deck, a 400's deck and a 600's deck. The order in which these decks were read was different for each subject (see Table B in the Appendix).

In all, there were 40 test cards in the 50's deck, 48 in the 100's deck, 36 in the 200's deck, and 24 in each of the 400's and 600's decks. Within each deck, dial types and sizes were rotated in systematic fashion.

Each day's readings on a test deck was preceded by a short warm-up run on a series of practice cards. Test sessions were approximately one hour in length. They were interrupted by short, two-minute rest periods at the conclusion of the warm-up readings and after each quarter of the test readings.

The subjects were instructed to read the dials as rapidly as possible and to make each reading to the nearest unit. Accuracy of reading was not mentioned as such in the instructions, but emphasis was placed on "reading to the nearest unit." Instructions were reviewed before each test session.

---

*These dial materials were prepared prior to the publication of the report of the Army-Navy-NRC Vision Committee, Subcommittee on standards to be employed in research on Visual Displays, in which standardized dial markings were recommended for use in experiments on instruments which might be used under night illumination conditions. (Army-Navy-NRC Vision Committee, 1947). The present 2.8 inch dials approximate the proposed standards, deviating most with respect to thickness of the heaviest graduation marks and with respect to pointer shape.

VI. HANDLING OF THE DATA

The reading errors recorded in the course of the experiment are classified and summarized in Table I.

Previous reports (Kappauf, Smith, and Bray, 1947; Kappauf and Smith, 1950) have discussed the occurrence of errors which are easily identified as involving the misreading of a scale number which is partially concealed by the pointer. Errors of this sort occurred in the present experiment in spite of attempts to eliminate them. Because their origin is known, and because they are not associated with reading the scale as such, these errors, 73 in all, are not considered in the discussion which follows. The data which are important for this report are those given in the rows of Table I beginning with the legend, "Total incorrect scale readings."

TABLE I

GENERAL BREAKDOWN OF READING ERRORS

Dials:	50's	100's	200's	400's	600's
Total readings made	8000	9600	7200	4800	4800
Total incorrect readings	357	526	1319	2203	2776
Errors of number identification		11	20	42	
Total incorrect scale readings	357	515	1299	2161	2776
Errors of 1 or 2 units	296	463	1114	2040	2600
Errors of 3 or 4 units	1	3	25	27	61
Errors of 5 or more units	62	54	193	173	214

The distribution of scale reading errors observed for each type and size of dial was a composite of what may be variously described as local, rounding, interpolation, or random errors on the one hand, and larger, systematic errors on the other. These two classes of errors differ in origin as well as in seriousness of consequence in practical situations, and it is therefore appropriate to consider them separately when assessing dial design variables. Various procedures might be followed to accomplish their separation (e.g. Ford, 1949; Horton 1949a), but in the present case the most informative comparisons between dial types result when the errors are simply classified by size. It is of practical importance to know, for example, the relative number of errors of two units on dials graduated by tens and by two unit steps, even though in the one case the errors are in interpolation and in the other they may represent an incorrect assessment of the subdivision values. Accordingly, in the discussion which follows, errors are treated in the two categories, "local errors" and "errors of five or more units." Errors tallied in the first of these categories are the errors of one and two units. The percentage of readings in error by these amounts is taken as an index of the precision of local scale reading for each type and size of dial. The tally of errors of five or more units may be interpreted as a measure of the likelihood of occurrence of systematic errors on each dial. Typical errors in this category were the errors of 5, 10, 20, 50 and various hundreds of units. Although this treatment of the observed errors gives no direct consideration to errors of three and four units, it is not difficult to show that conclusions from the study are little influenced by the specific error breakdown employed. (See Tables II and III below).

The data in Table I deserve one further comment. This is that the total number of incorrect scale readings is regularly smaller than the sum of the errors in the sub-categories. This arises from the fact that whenever readings clearly involved both local and gross errors (e.g. errors of + 101, -19, + 52) they were tallied both in the "one or two unit" category and in the "five units or larger" category.

VII. RESULTS

Frequency of Errors of One and Two Units

The percentage of readings in error by one or two units is shown for the various types and sizes of dials in Figures 2 to 4. Figures 2 and 3 deal separately with dials of the two sizes, and within each size group permit direct observation of the effect of scale range and graduation scheme. Figure 4 reproduces the same data, but classifies dial groups by graduation scheme and brings together dials of the two sizes.

In all of these figures, error frequencies are plotted as a function of the size of the called interval (terminology of Garner and Gebhard, 1949). Since the subjects in this experiment always read or "called" to the nearest unit, size of the called interval was for each dial equal to the length of arc devoted to a single scale unit on that dial. This treatment of the data in terms of size of called interval is in accordance with the results of recent work which has shown that the degree to which a scale is expanded or compressed is of central importance in the determination of interpolation errors. Dial size and scale range have effects on interpolation errors (best seen in Tables II and III below) but they enter the present analysis only indirectly as factors which influence the size of called interval.

In Figures 2 and 3, the scale of called interval is logarithmic. The percentage error scale on the other hand follows the angular transformation. Use of this transformed percentage scale has two advantages. The first and primary one is that it makes the variance of the observed percentages roughly the same in terms of units on the transformed scale. The second and incidental advantage is that it spreads out the small percentage values where dial to dial differences are small.

Figures 2 and 3 show the extent to which local errors are reduced when any given scale is graduated by five unit intervals instead of by tens, and how they are further reduced when that scale is graduated by units or twos instead of by fives. Differences between the error frequencies for dials graduated by units and by twos are not significant when tested by Chi Square, but it is probable that this particular result was to some extent influenced by the fact that in this study the dial pointers were always set close to exact unit positions (see page 4 above). Units graduation might prove more satisfactory than graduation by twos under more general test conditions.

Shown also, and particularly in Figure 2, is the manner in which finer graduation schemes progressively lose their advantage when size of the called interval becomes too small. Thus the present data suggest that, so far as local scale reading errors are concerned, graduation by fives loses its advantage over graduation by tens at a size of called interval of about 0.007 inches, and that graduation by twos loses its advantage over graduation by fives at a called interval size of about 0.022 inches.

Data obtained with the 1.4 inch dials (see curves based on solid symbols in Figure 4) show that the error frequencies decline sharply as called interval increases to 0.044 inches, and that beyond this point, error frequencies are essentially stable. The only 1.4 inch dials which showed a statistically significant error decrease when called interval changed from 0.044 inches to 0.088 inches were the dials graduated by tens, but even for these dials the amount of the decrease was small. Taken

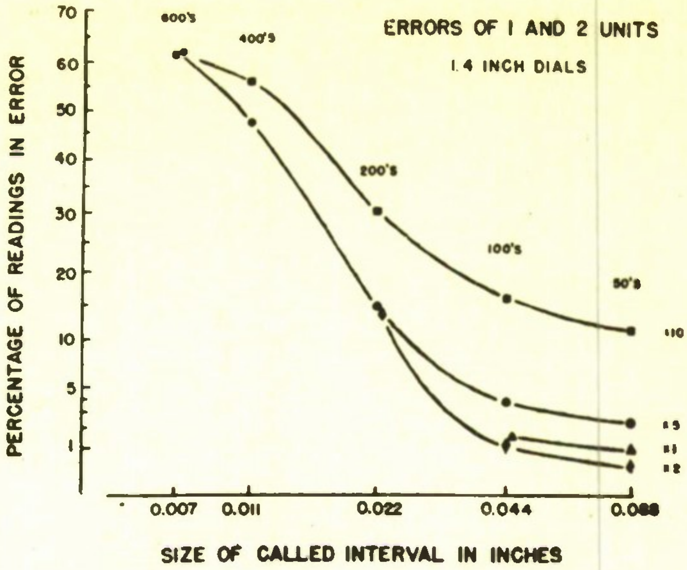

Figure 2. Frequency of reading errors of one and two units on all dials of the 1.4 inch size: graduation scheme as parameter.

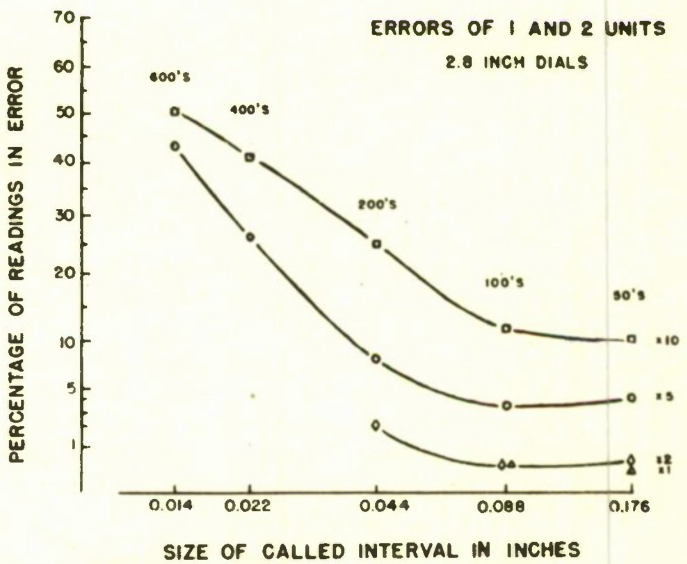

Figure 3. Frequency of reading errors of one and two units on all dials of the 2.8 inch size: graduation scheme as parameter.

Note: In these figures, the called interval scale is marked off logarithmically. The scale of percentage wrong readings has been transformed under the angular transformation. This has the advantage of making the variance of each of the observed error percentages roughly the same in terms of units on the transformed scale. It also spreads out the small percentage values where differences become small.

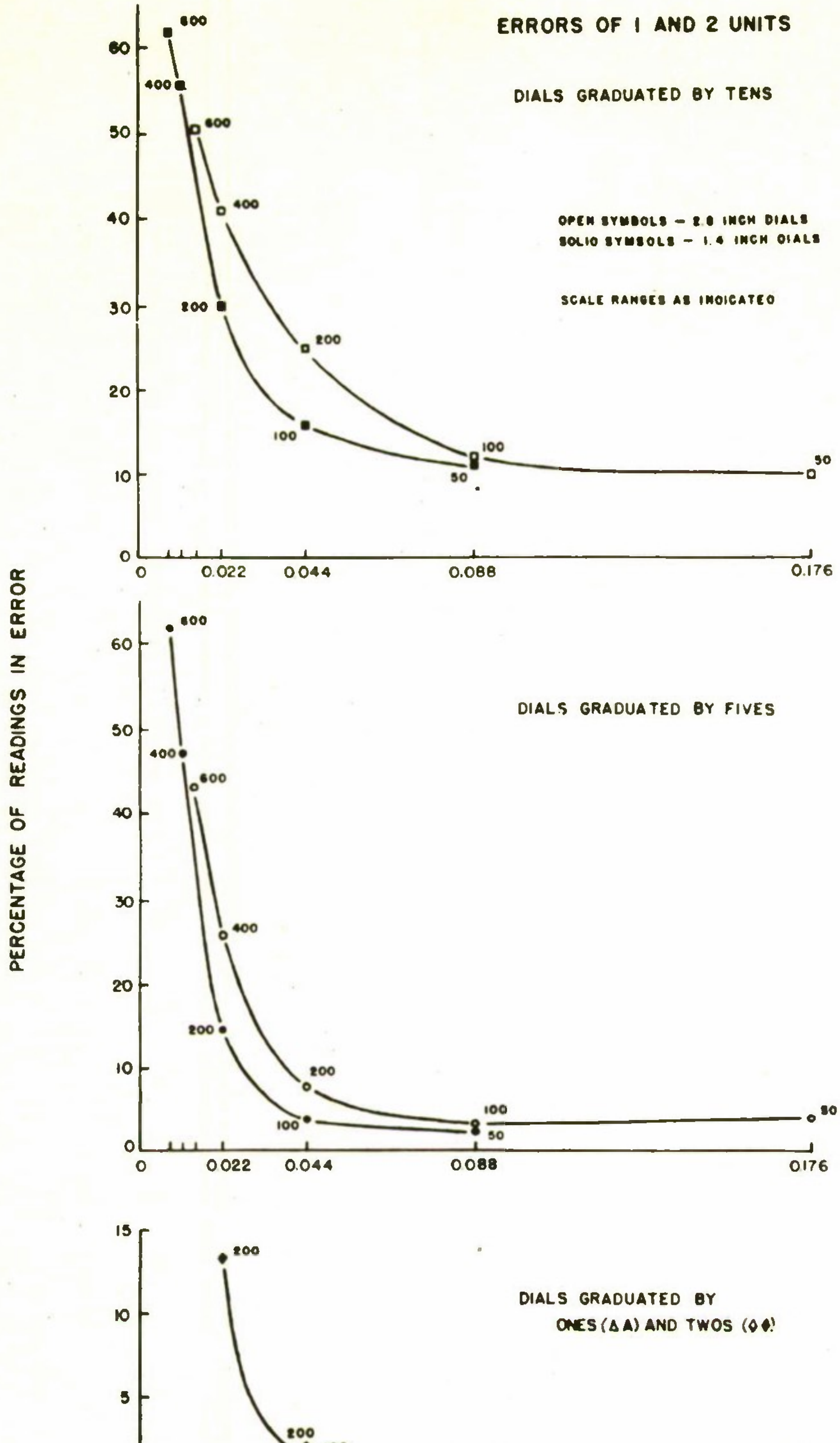

Figure 4. Frequency of reading errors of one and two units for all dials: Errors compared by dial size.

Same data as in Figure 3, but plotted on linear scales of called interval and percentage errors. Note expanded ordinate scale on graph for dials graduated by ones and twos.

together, the curves suggest that local errors in reading the 1.4 inch dials reach a relatively stable minimum when the size of the called interval has become about 0.05 inches, and that this critical size is influenced only slightly by the nature of the graduation scheme employed.

These data for the 1.4 inch dials fit in well with the results of several other recently reported experiments which have inquired into the magnitude and frequency of interpolation errors as a function of called interval. These studies by Grether and Williams (1947, also 1949), Leyzorek (1949) and Horton (1949b) used scales having graduation marks or radar scope range rings of approximately the same thickness as the marks on the present 1.4 inch dials, and it is probably for this reason that the agreement in results is so good. Each of these authors has concluded that interpolation errors are minimized when the size of the marked interval is 0.5 to 0.6 inches or larger. Inasmuch as interpolation to tenths was explicitly called for by Grether and Williams and by Horton, and was probably the task which most subjects set for themselves in Leyzorek's study, these data indicate that the critical size of called interval for scales read to tenths of marked intervals is about 0.05 or 0.06 inches. The present records for the 1.4 inch dials bear out this general conclusion and extend it by showing that the critical size of called interval for scales graduated more finely than by tens is also about 0.05 inches.

Data obtained with the 2.8 inch dials (see curves based on open symbols in Figure 4), imply that minimum error frequencies are not approached until size of called interval is nearer 0.088 inches. There is a general displacement of the curves for these data, away from the curves based on the small dial data, in the direction of less precise readings. At called interval sizes of both 0.022 inches and 0.044 inches on dials graduated by tens as well as on dials graduated by fives, there is a statistically significant difference between the number of reading errors made on the large and small dials (Chi Square Test). In view of the Grether and Williams evidence that dial size as such has no effect on interpolation errors when graduation mark thickness and pointer thickness are held uniform on all dials, the present difference in error data from large and small dials with identical size of called interval is interpreted to be the result of differences in the fineness of graduation marks, and perhaps in the mass of the pointers, on the two sets of dials used here. A partial check of errors made in the vicinity of heavy and light graduation marks supports this suggestion.

Frequency of Errors of Five or More Units

The percentage of readings which were in error by five or more units is listed for each type and size of dial in part A of Table II below and is shown graphically in Figure 5. In this figure the scale of percentages is again arranged in accordance with the angular transformation.

Inspection of Figure 5 reveals no general or systematic effect of dial size or of graduation scheme on the frequency of these larger errors. Tests of homogeneity, using Chi Square or binomial probability paper (Mosteller and Tukey, 1949), lead one to accept the hypothesis that dial size and graduation scheme are without effect for all sets of dials except the 600's dials. Among the latter, those graduated by fives were read with significantly more errors than those graduated by tens. This difference for the 600's dials, incidentally, did not arise because of any one particular kind of large error, for errors of every classifiable sort were more frequent on the by fives dials than on the by tens. It is probably important that the 600 scale graduated by fives was the most crowded dial used in the experiment. This makes it appear that the proper conclusion from these data is to the effect that graduation scheme per se has no direct effect upon the frequency of large scale reading errors (as classified here) unless the size of the marked interval becomes too small.

Scale range, on the other hand, has an important effect on these errors. The 200's, 400's, and 600's dials were read with more errors than the 100's or 50's dials. The difference between the average error rate for all the 100's dials and that for

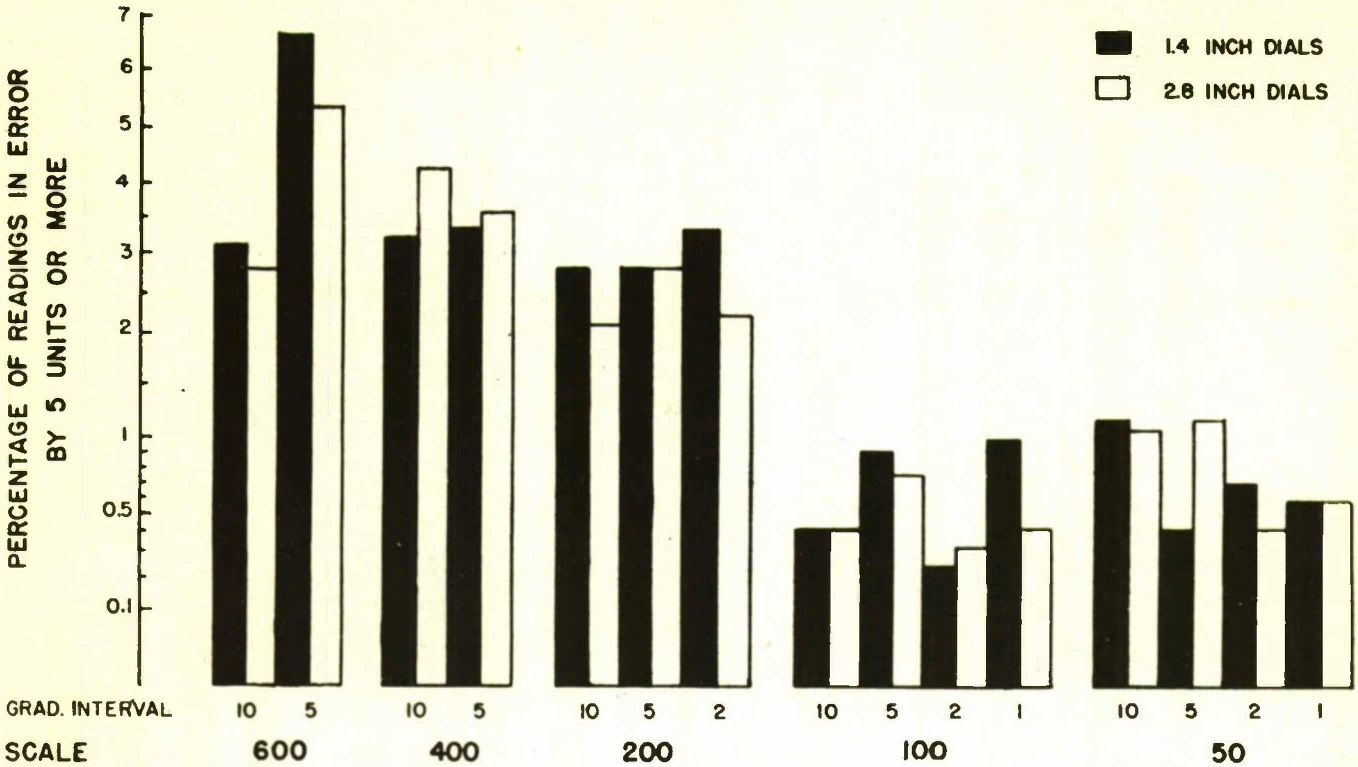

Figure 5. Frequency of reading errors of five or more units for all dials.

all the 50's dials is not statistically significant, but all the succeeding differences, involving an increase in the average frequency of error with increasing scale range, are significant. It should be noted, however, that the significant increase in error rate with change in scale range from 400 to 600 is not unrelated to the already mentioned large number of errors on the 600's dials graduated by fives.

Among possible explanations for the general increase in systematic errors with scale range one may quickly discount suggestions involving change in called interval (which is without effect when dial size varies) or the reading of hundreds digits on the scales of longer range (which proves not to account for many errors). The factors which appear to be of importance are the changes in scale numbering and in the number of principal subdivisions between scale numbers which are a necessary consequence of increasing scale range on a dial of a given size. Thus, in reading a 600's dial, one has to identify both the tens digit and the final digit by scale inspection, whereas in reading a 50's or 100's dial he need only find the final digit by scale inspection since the tens digit is given in the scale numbering. The scales might be said to differ in complexity. The frequency of large reading errors increases with this complexity, or with the amount of work the reader has to do in determining the scale values of principal subdivision marks nearest to the pointer.

General Tabulation of Error Data

For the reader who is interested specifically in error magnitudes or their distribution for each of the dials used, Tables II and III (pages 12 and 13) have been prepared. These tables cover the same data which are already summarized in Figures 2 to 5, but present them in a form which is more usable for certain purposes. For the design engineer, these tables permit a direct answer to the question, "How well did the present group of 20 subjects perform in terms of an error tolerance of 'x' units or 'y' percent of full scale?" For those concerned with reading error distributions as such, Table II permits a reconstruction of the distribution of readings which were in error by 0, 1, 2, and 3 or 4 units.

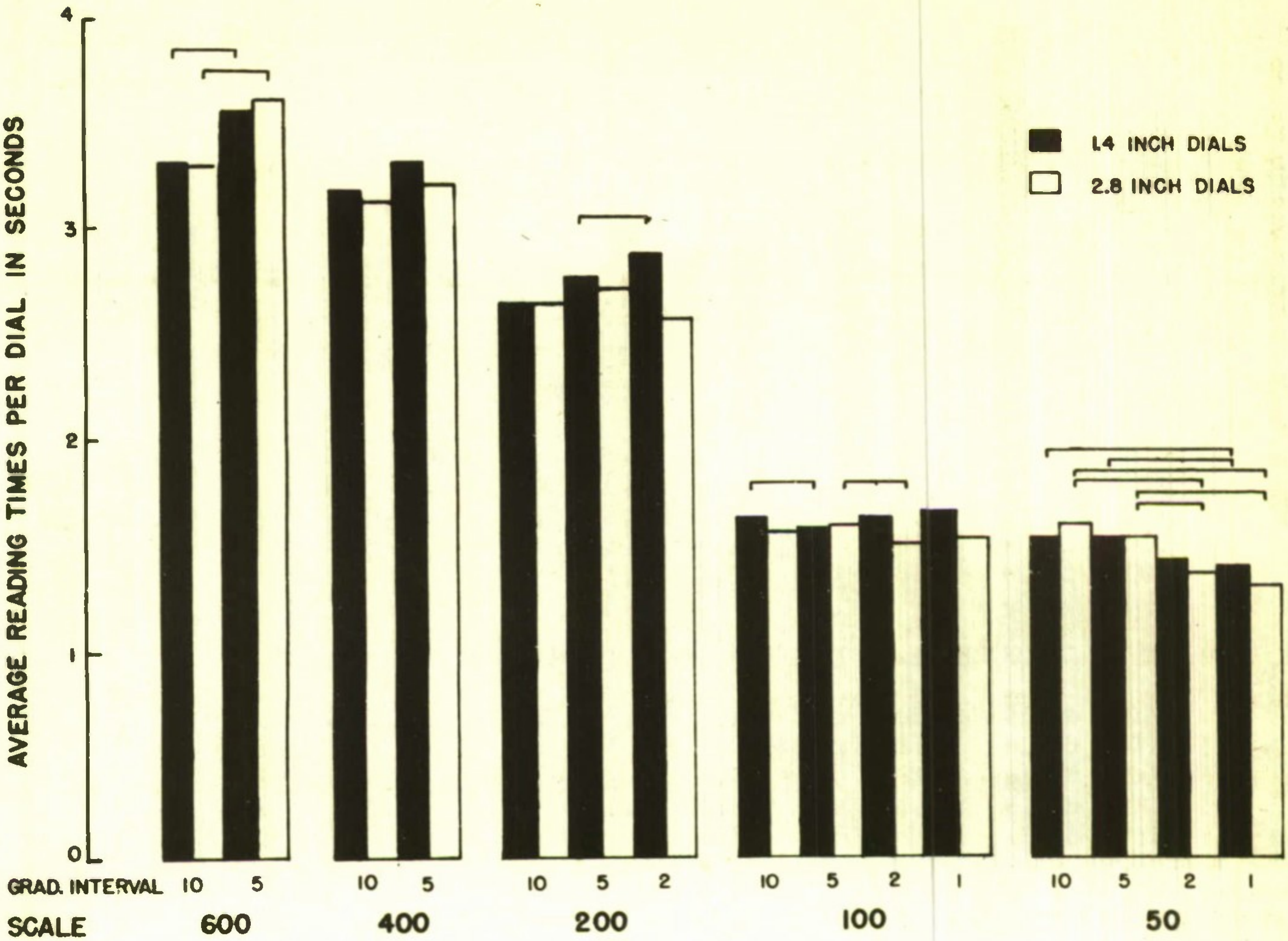

Figure 6. Reading times for all dials. For explanation of brackets, see text.

Reading Times

Reading times are presented in Figure 6, and in Table IV on page 13.

The principal results in regard to reading time closely resemble those just outlined in regard to the incidence of large errors: namely, that dial size and graduation scheme within the limits studied have little effect in comparison with the effect of scale range.

Over the entire experiment, the effect of dial size, though small, is statistically reliable in the direction of faster reading for the 2.8 inch dials. This confirms the data of two earlier studies using similar experimental equipment and test procedures (Kappauf, Smith, and Bray, 1947; Kappauf and Smith, 1948).

Crude tests of the statistical significance of the differences in reading times between graduation schemes were made by counting the number of subjects who read any one dial in better average time than each other dial of comparable size and scale range. Those differences which met the criterion that they reflected the behavior of at least 16 of the 20 subjects are marked by the small brackets in Figure 6. Of greater interest than these local differences, however, is the general trend from the situation met with the 50's dials where the use of finer graduation schemes

TABLE II

FREQUENCY OF ERRORS GREATER THAN A STATED NUMBER OF SCALE UNITS

	No. readings for each type and size	2.8 inch diam. dials graduated by				1.4 inch diam. dials graduated by			
		10	5	2	1	10	5	2	1
A. PERCENTAGE OF READINGS IN ERROR BY 5 OR MORE UNITS									
50's dials	1000	1.1	1.2	0.4	0.6	1.2	0.4	0.7	0.6
100's dials	1200	0.4	0.8	0.3	0.4	0.4	0.9	0.2	1.0
200's dials	1200	2.1	2.8	2.2		2.8	2.8	3.3	
400's dials	1200	4.3	3.6			3.2	3.3		
600's dials	1200	2.8	5.3			3.1	6.6		
B. PERCENTAGE OF READINGS IN ERROR BY MORE THAN 2 UNITS									
50's dials	1000	1.1	1.3	0.4	0.6	1.2	0.4	0.7	0.6
100's dials	1200	0.4	0.8	0.3	0.5	0.4	1.0	0.3	1.0
200's dials	1200	3.1	2.9	2.4		3.4	2.9	3.4	
400's dials	1200	4.9	3.6			4.3	3.8		
600's dials	1200	3.4	5.5			4.7	9.6		
C. PERCENTAGE OF READINGS IN ERROR BY MORE THAN 1 UNIT									
50's dials	1000	1.1	1.3	0.9	0.6	1.4	0.4	1.1	0.8
100's dials	1200	0.6	0.8	0.7	0.5	0.7	1.0	0.9	1.0
200's dials	1200	5.1	3.0	3.8		5.4	3.0	5.2	
400's dials	1200	8.2	5.2			12.1	8.9		
600's dials	1200	8.5	9.2			16.5	24.0		
D. TOTAL PERCENTAGE OF READINGS IN ERROR									
50's dials	1000	10.9	5.4	0.9	0.8	12.3	2.7	1.1	1.6
100's dials	1200	12.2	4.1	0.8	0.9	16.2	4.7	1.4	2.6
200's dials	1200	27.4	10.1	4.2		32.5	17.5	16.6	
400's dials	1200	43.7	28.5			58.2	49.8		
600's dials	1200	52.2	46.6			64.2	68.2		

TABLE III

FREQUENCY OF ERRORS GREATER THAN A STATED PER CENT OF FULL SCALE

	No. readings for each type and size	2.8 inch diam. dials graduated by				1.4 inch diam. dials graduated by			
		10	5	2	1	10	5	2	1
A. PERCENTAGE OF READINGS IN ERROR BY MORE THAN 2% OF FULL SCALE									
50's dials	1000	1.1	1.3	0.9	0.6	1.4	0.4	1.1	0.8
100's dials	1200	0.4	0.8	0.3	0.5	0.4	1.0	0.3	1.0
200's dials	1200	2.1	2.8	2.2		2.8	2.8	3.3	
400's dials	1200	4.0	3.1			3.0	2.6		
600's dials	1200	1.3	2.6			1.4	2.0		
B. PERCENTAGE OF READINGS IN ERROR BY MORE THAN 1% OF FULL SCALE									
50's dials	1000	Only read to nearest 2%							
100's dials	1200	0.6	0.8	0.7	0.5	0.7	1.0	0.9	1.0
200's dials	1200	3.1	2.9	2.4		3.4	2.9	3.4	
400's dials	1200	4.3	3.6			3.2	3.3		
600's dials	1200	2.6	5.3			2.9	6.1		
C. PERCENTAGE OF READINGS IN ERROR BY MORE THAN 0.5% OF FULL SCALE									
50's dials	1000	Only read to nearest 2%							
100's dials	1200	Only read to nearest 1%							
200's dials	1200	5.1	3.0	3.8		5.4	3.0	5.2	
400's dials	1200	4.9	3.6			4.3	3.8		
600's dials	1200	2.7	5.3			3.4	7.5		

TABLE IV

AVERAGE READING TIMES PER DIAL - IN SECONDS

	No. readings for each type and size	2.8 inch diam. dials graduated by				1.4 inch diam. dials graduated by			
		10	5	2	1	10	5	2	1
50's dials	1000	1.62	1.55	1.39	1.34	1.55	1.55	1.45	1.42
100's dials	1200	1.59	1.61	1.52	1.55	1.65	1.60	1.65	1.68
200's dials	1200	2.65	2.72	2.58		2.65	2.78	2.88	
400's dials	1200	3.14	3.21			3.19	3.32		
600's dials	1200	3.30	3.63			3.32	3.56		

speeds up reading, to the situation at the longest scale range where finer graduation causes reading to be slower.

Reading speed differences between dials of different scale range, but similar in size and graduation scheme, were checked by the same subject counting test outlined above. In these terms, all differences between the 400's and 200's dials and between the 200's and 100's dials were significant, along with those differences between the 100's and 50's dials graduated by units and that between the small 100's and 50's dials graduated by twos.

In the routine of the experiment, it will be remembered, all dials of a given scale range were read by any one subject in a single test session. This plan favored a desired familiarity with the number scales on each set of dials and kept confusion between dials at a minimum, but may have given subjects a set to use a rather general reading pace for each particular session, a pace maintained within limits in spite of changes in scale graduation and dial size. The possibility to be considered then is whether the foregoing reading time differences between dial sizes and graduation schemes are spuriously small. As a test of this point, at least as regards graduation schemes, four college students were run in a short supplementary experiment in which each day's reading included cards of large 100's, 200's, and 600's dials presented in various random sequences. As in the main experiment, the subjects were always advised of the type of dial they would read on the next card, and all dials on a given card were of identical design. The results of the five day tests reproduced very closely the effect of scale range observed in the main experiment. In fact, relative differences in reading times between the 100's, 200's, and 600's dials were almost identical with those shown in Figure 6. Time differences between graduation schemes were, for three of the subjects, irregular and typical of the data for the 20 subjects of the main experiment. The fourth subject was consistently faster in reading the more finely graduated 100's and 200's dials. The average data, as influenced by this one subject, exhibited a progressive effect of graduation scheme comparable to the relative speed changes observed for the 2.8 inch 50's dials in the main experiment (see Table IV).

These data suggest that the original test conditions did not seriously distort the reading time records for different graduation schemes. The greatest decrease in reading time to be expected in going from graduations by ten unit intervals to graduations by units when dealing with dials of short scale range is of the order of 15 to 20 per cent. This is to be compared with a reduction of about 50% in reading time in going from 600's dials to 100's or 50's dials. The conclusion stated above may therefore be repeated, namely, that the principal variables affecting dial reading time are scale range and such factors as number scheme and the hierarchy of subdivision markings which necessarily change with scale range. Of secondary importance in influencing reading time are dial size and scale graduation scheme.

Summary Interpretation

The preceding paragraphs have presented data describing the effects which certain graduation schemes, scale ranges and dial sizes have upon (1) the precision of local scale reading, (2) the occurrence of large or systematic scale reading errors, and (3) the speed of reading. Precision of reading varies significantly with size of the called interval and with graduation scheme, and seems to depend secondarily upon graduation mark thickness. The frequency of large errors and the speed of reading, on the other hand, depend primarily upon the complexity of the scale. Hence they vary critically with scale range, but are influenced much less by dial size (within the limits studied) or by the specific forms of graduation scheme employed.

VIII. DISCUSSION

Results obtained under different Reading Attitudes

In an earlier study of this series (Kappauf and Smith, 1948), data were obtained on the dial reading performance of a group of six subjects who were instructed to read as accurately as possible. The stimulus materials used at that time were chosen from those then in preparation for the present study. Although a number of variables in that preliminary experiment were admittedly not too well controlled and not very many readings were involved, trends in the data check out well with the present results. As in the present study, reading time was influenced most by scale range. Total errors (a composite of local and systematic errors, with local errors predominating) were dependent upon the size of the called interval and upon graduation scheme. The principal differences between the two sets of data may be summarized roughly by saying that the accuracy-instructed subjects made about 30% fewer reading errors, while using reading times which were about 30% longer than those reported here for speed-instructed subjects. This observation of a 30% decline in errors of quantitative reading with change of instructions agrees well with the data of another experiment in which a single group of subjects read with different reading attitudes during different test sessions (Kappauf, Smith, and Bray, 1947).

Studies on Interpolation Accuracy vs. Size of Called Interval

Taken together, the results of Grether and Williams (1947, also 1949) and of Leyzorek (1949) and those of the present study for dials of the 1.4 inch size indicate that the accuracy of interpolation between marks of a given thickness is dependent upon the separation of those marks and is independent of essentially all other factors in the design of the display -- the size of the display, the arrangement of markers, the form of the display, etc. In spite of minor differences between these studies in experimental technique and in the indices of reading precision which were adopted, evidence is consistent that the accuracy of visual interpolation to tenths is at its best when mark separation is about 0.5 inches or more. This result applies principally for reading distances of 24 to 30 inches, but may hold for distances as short as 12. It applies to scales where the marks are about 0.015 inches in stroke thickness.

One problem still in need of investigation to round out our data on scale design in relation to interpolation accuracy is that of determining the extent to which graduation mark thickness influences reading errors. Results on this subject will be of interest, not in regard to aircraft instrument design, but in regard to the design and use of slide rules, computers, and various precision instruments where small marked intervals and fine graduation markings have traditionally been used.

Studies on the relative merits of different Graduation Schemes when Called Interval remains Constant

Previous studies dealing with the question of the relative merits of scales with different minor graduation marks have been reported by Loucks (1944) and Vernon (1946). Vernon dealt with the particular case where readings are made to the nearest marked interval. Her interest was in scale interpretation errors, and although her results show that the easiest scale divisions to read are, in order, divisions with values of one, two, and five units (or tens multiples of these), they do not relate directly to the results of the present experiment. Vernon's criterion of evaluation was different, and the units in terms of which her readings were called differed from scale to scale. A portion of the work of Loucks, however, did bear on the same aspect of the graduation scheme problem as the present experiment.

Loucks' observations were made as part of an investigation of possible improvements in the design of tachometer dials. He compared, among others, the three large dials shown in Figure 7. Major divisions on these dials are 100 unit steps. The graduation intervals on the three dials are 20 units, 50 units, and 100 units respectively. Loucks presented these dials in tachistoscopic exposure, using both 1.5 second and 0.75 second exposure intervals. He instructed his subjects to report their readings to the nearest 20 RPM and scored readings in error when they failed to do so. The percentage of readings recorded as errors is shown under each of the dial photographs in the figure. The dial graduated by 20 unit intervals was read with significantly more errors at both exposure times than was the dial graduated by 100 unit intervals. The latter was also significantly better than the dial graduated by 50 unit intervals when the 0.75 second exposure time was used.

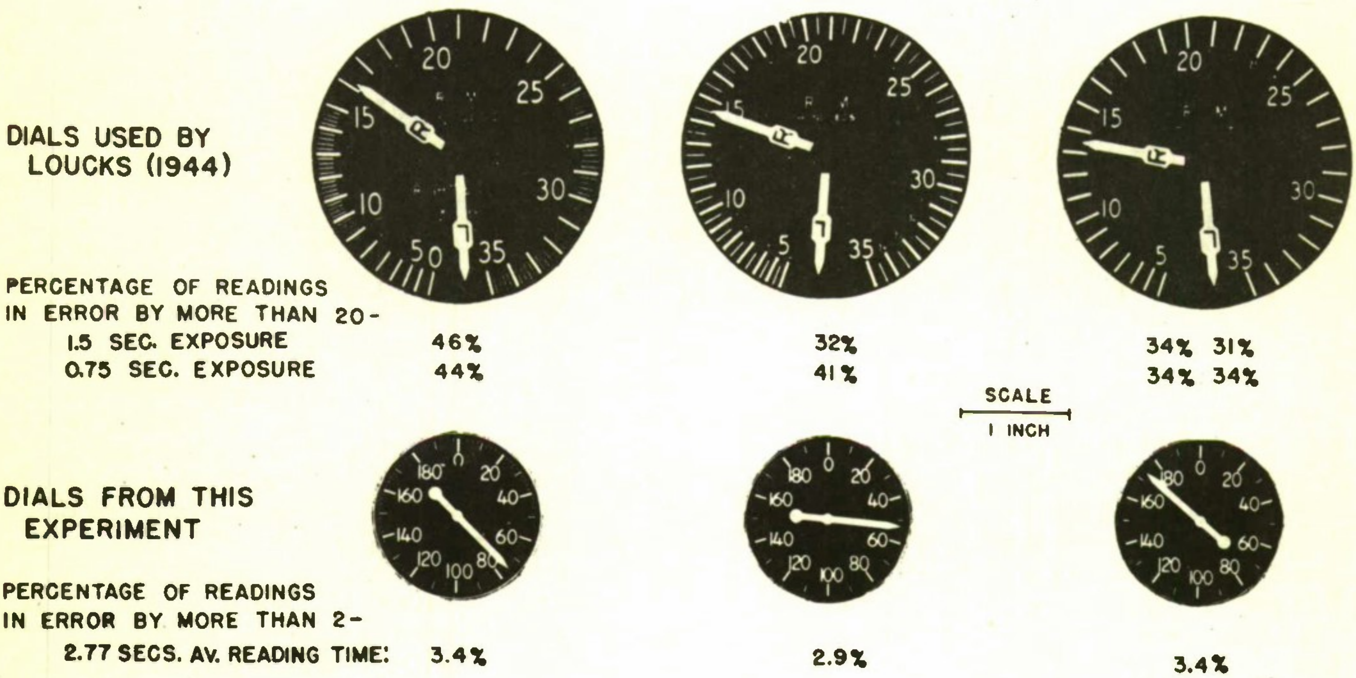

Figure 7. Comparison with dials and data from Loucks (1944).

These results are properly interpreted to mean that relatively simple scaling, without fine subdivisions, is desirable on dials which must be read in brief exposures. An uncautious reader, however, who does not regard these reading conditions too seriously might infer that fine subdivisions are undesirable in any situation where quantitative scale readings must be made as rapidly as possible. For this reason it is interesting to compare Loucks' data with the records of the present experiment for the dials which most closely resemble his. It turns out that the scales on Loucks' tachometer dials correspond almost exactly in graduation mark spacing to the scales on the 1.4 inch 200's dials. Figure 7 permits their direct comparison. An error tolerance of two units on these scales graduated by twos, fives, and tens, is comparable to the 20 RPM tolerance used on the tachometer dials. So Table II above has been consulted to determine for these 200's dials the percentage of readings in error by more than two units. The percentages have been copied into Figures 7 and appear below the three dials. These error frequencies are all small and do not differ significantly from each other. The two sets of data, therefore, are quite different.

Part of this difference may have arisen because of differences in the relative "cleanness" of the dials in the two experiments. Loucks, in discussing his results, attributed poor performance on the more finely graduated dials to their cluttered, busy appearance. His conclusion, based on additional later work, was that for optimum legibility, scales should be kept as clean as possible. It might be argued, then, that the present 200's dials graduated by twos and fives may have had some advantage over Loucks' dials in having shorter, more easily differentiated markings.

But more important in accounting for the difference in results is the difference between the allowed reading times in the two experiments. Loucks used brief tachistoscopic exposures. The present subjects read at their own pace. In so doing they took two to four times as long as Loucks' subjects were given, and would have taken longer still to read dials numbered like the tachometers (see reading time discussion above). In other words, Loucks' exposure times were considerably shorter than men use when reading as rapidly as possible, and his results should not be expected to apply generally to all hurried reading situations. That such a change in procedure might be significant in regard to the results obtained was recognized by Loucks when he predicted early in his paper that, "The dial which gives the fewest errors for brief exposure readings is not necessarily the one which will make possible the most precise readings were the subjects to have unlimited time." Figure 7 bears him out, and indicates a general need for the cautious interpretation of experimental results in terms of test procedures employed in collecting the data. In view of Horton's observations (1949b) to the effect that certain trends in radar scope reading errors are uninfluenced by using tachistoscopic exposures of three and five seconds as compared with unrestricted time of exposure, it may be that the experimental result which needs most careful scrutiny before it is generalized is that which is based upon a test procedure using exposures under three seconds.

So far as the specific experiments compared in Figure 7 are concerned, it seems clear then that they should be taken as supplementary, rather than contradictory. As such, they show that choice of graduation scheme may vary in accordance with the task at hand (in so far as it dictates reading speed or stress), with the magnitude of the tolerable reading error (see Table II above), and with the size of the called interval (see Figures 2 and 3 above). Except for very short reading times, shorter than subjects take when instructed to read to the nearest unit as rapidly as possible, finer graduation schemes are to be preferred in quantitative reading because they serve to reduce the frequency of local errors of interpolation, rounding and the like. Mark crowding limits the utility of extra minor graduation markings for reducing local errors, but as shown in Figure 2, there is a considerable range over which graduation by fives is preferable to graduation by tens and another over which graduation by twos or units is still more satisfactory than graduation by fives.

The Use of Standardized Instrument Scales reading in Per Cent

The evidence presented in this paper has pointed to the close similarity between reading performance on the 100's and 50's dials which were used. These two sets of dials were similar in number scheme, both being numbered by tens. Seemingly both sets of dials can be read to the nearest unit with essentially maximum speed and accuracy. These observations are encouraging in regard to one proposed scheme for standardizing certain aircraft instrument displays, namely, by transforming their scales so that all will read in per cent of normal or maximally effective operating conditions. Such instruments scaled from 0 to 100 (or say 120) and numbered by tens would certainly be read very well.

IX. CONCLUSIONS

Local or small errors in quantitative scale reading vary with the size of the called interval and with graduation scheme. They seem also to be influenced by graduation mark thickness, a matter which deserves special study in and of itself.

Systematic or large scale reading errors, as well as reading times, are determined primarily by scale range and the factors of scale organization which necessarily change with scale range. They are influenced only secondarily by dial size and graduation scheme.

The reading speed and error data for the 100's dials are such as to lend strong support to the suggestion that particular aircraft instruments be rescaled to indicate per cent of normal operating condition.

Procedural differences between various scale legibility experiments may influence the trends observed within the data. For this reason test data should regularly be interpreted in terms of test conditions and the degree to which the latter resemble conditions of the job to which the data would be applied.

X. BIBLIOGRAPHICAL REFERENCES

Chapanis, A. Accuracy of interpolation between scale markers as a function of scale interval number. Amer. Psychologist. 1947, 2. 346. (Unclassified, English)

Ford, A. Types of errors in location judgments on scaled surfaces. II. Random and systematic errors. J. appl. Psychol. 1949, 33. pp 382-394. (Unclassified, English)

Ford, A., and M. H. Getz. Types of errors in reading GCA scaled scopes. Technical Report No. 4, Contract W28-099-ac-130, Watson Laboratories, AMC, USAF. August 31, 1948. (Restricted, English)

Garner, W. R., and J. W. Gebhard. Machine Dials and Scales. Machine Design. August 1949. (Unclassified, English)

Grether, W. F., and A. C. Williams. Speed and accuracy of dial reading as a function of dial diameter and spacing of scale divisions. US AAF AMC, Engng. Div., Aero Med. Lab., TSEAA-694-1E. March 31, 1947. p 22. (Unclassified, English)

Grether, W. F., and A. C. Williams, Jr. Psychological factors in instrument reading: II. The accuracy of pointer position interpolation as a function of the distance between scale marks and illumination. J. appl. Psychol. 1949, 33. pp 594-604. (Unclassified, English)

Horton, G. P. An analysis of errors made on a schematic PPI display. US AAF AMC, Engng. Div., Aero Med. Lab., Technical Report No. 5960. October 1949a. (Unclassified, English)

Horton, G. P. Accuracy of reading target location and size of schematic PPI display. US AAF AMC, Engng. Div., Aero Med. Lab., Technical Report No. 5961. October 1949b. (Unclassified, English)

Kappauf, W. E., and W. M. Smith. Design of instrument dials for maximum legibility: II. A preliminary experiment on dial size and graduation. US AAF AMC, Engng. Div., Aero Med. Lab., MCREXD-694-1-N. July 12, 1948. p 16. (Unclassified, English)

Kappauf, W. E., and W. M. Smith. Design of instrument dials for maximum legibility: III. Some data on the difficulty of quantitative reading in different parts of a dial. US AAF AMC, Engng. Div., Aero Med. Lab., Technical Report No. 5914. January 1950. p 13. (Unclassified, English)

Kappauf, W. E., W. M. Smith, and C. W. Bray. Design of instrument dials for maximum legibility: I. Development of methodology and some preliminary results. US AAF AMC, Engng. Div., Aero Med. Lab., TSEAA-694-1L. October 20, 1947. p 42. (Unclassified, English)

Lebensohn, J. E. Scientific and practical considerations involved in the near-vision test with presentation of a practical and informative near-vision chart. Amer. J. of Opthalmology. 1936, 19. pp 110-117. (Unclassified, English)

Leyzorek, M. Accuracy of visual interpolation between circular scale markers as a function of the separation between markers. J. exper. Psychol. 1949, 39. pp 270-279. (Unclassified, English)

Loucks, R. B. Legibility of aircraft instrument dials: The relative legibility of tachometer dials. US AAF, School of Aviation Medicine, Randolph Field, Texas, Project No. 265, Report No. 1. May 30, 1944. p 8. (Unclassified, English)

Mosteller, F., and J. W. Tukey. The uses and usefulness of binomial probability paper. J. of Amer. Stat. Assn. 1949, 44. pp 174-212. (Unclassified, English)

U. S. Army-Navy-NRC Vision Committee. Report by subcommittee of the Army-Navy-NRC Committee on standards to be employed in research on visual displays. Minutes and proceedings of the 19th meeting. May 27-28, 1947. pp 13-18. (Restricted, English)

Vernon, M. D. Scale and dial reading. Unit in Applied Psychology, the Psychological Laboratory, Cambridge, A.P.U. 49. June 1946. (Unclassified, English)

TABLE A

DESCRIPTIVE DATA ON SUBJECT POPULATION

SCORES OBTAINED ON SUB-TESTS OF THE ARMY AIR FORCE NORMATIVE TEST BATTERY

SUBJECT	AGE	SCHOOL GRADE	HENMON NELSON TEST SCORE (avail. thru H.S.)	Graph Reading	Meter Reading	Table Reading	Number Filing	Number Size	Distance Perception	Numerical Operations	Vocabulary Level	Total Scores
D. T.	18	12	100	15	19	23	3	15	61	49	50	235
G. T.	16	11	101	10	0	20	11	26	58	51	47	223
J. Y.	18	12	101	6	-10	28	11	28	54	45	45	207
I. W.	18	12	102	15	8	20	6	16	82	53	45	245
C. A.	17	11	105	24	- 4	32	15	34	35	90	52	278
G. C.	17	12	109	17	20	28	19	34	38	61	48	265
F. B.	16	11	109	21	24	25	18	36	78	68	56	316
W. Bu.	15	11	110	5	23	27	16	30	30	56	52	239
R. C.	17	12	110	13	17	26	16	36	78	58	61	305
L. S.	17	11	110	4	2	15	12	29	44	45	54	205
T. C.	16	11	111	13	18	28	11	40	58	55	48	271
P. B.	16	11	118	22	36	31	29	44	68	120	51	401
R. D.	17	12	119	18	44	25	16	36	86	70	55	350
W. B.	17	12	119	15	35	34	12	44	49	128	65	382
W. Ba.	16	11	120	14	25	35	20	41	51	56	47	289
P. C.	15	11	121	23	17	28	9	22	45	36	58	238
R. S.	17	12	122	21	37	28	- 6	44	66	103	60	353
H. S.	17	11	123	23	14	31	16	34	54	78	55	305
F. C.	16	11	129	9	- 1	21	15	35	52	96	59	286
L. C.	15	11	135	21	37	29	15	35	54	51	66	311

APPENDIX

TABLE B

ORDER IN WHICH TEST DECKS WERE READ BY DIFFERENT SUBJECTS

Subject	Test Day 1	Test Day 2	Test Day 3	Test Day 4	Test Day 5
T. C.	50	100	400	200	600
I. W.	50	200	100	600	400
F. C.	50	400	600	100	200
P. B.	50	600	200	400	100
P. C.	100	200	600	400	50
W. Bu.	100	400	200	50	600
G. C.	100	600	50	200	400
W. B.	100	50	400	600	200
L. S.	200	400	50	600	100
D. T.	200	600	400	100	50
H. S.	200	50	100	400	600
R. S.	200	100	600	50	400
J. Y.	400	600	100	50	200
C. A.	400	50	600	200	100
L. C.	400	100	200	600	50
R. D.	400	200	50	100	600
G. T.	600	50	200	100	400
F. B.	600	100	50	400	200
W. Ba.	600	200	400	50	100
R. C.	600	400	100	200	50

TITLE: Design of Instrument Dials for Maximum Legibility - Part 4. Dial Graduation, Scale Range and Dial Size as Factors Affecting the Speed and Accuracy of Scale Reading - and Appendix (AF Technical Report)	ATI-Part of 78 314
AUTHOR(S) : Kappauf, William E.; Smith, William M.	DIVISION (None)
ORIG. AGENCY : Princeton University, N.J.	ORIG. AGENCY NO. (None)
PUBLISHED BY : AMC, Wright-Patterson Air Force Base, Dayton, O.	PUBLISHING AGENCY NO. AF-TR-5914, Part 4

DATE	U.S. CLASS.	COUNTRY	LANGUAGE	PAGES	ILLUSTRATIONS
Feb' 50	Unclass.	U.S.	English	21	tables, graphs

ABSTRACT:

The extent to which instrument-dial reading errors and reading times are influenced by the spacing of graduation marks, graduation-mark values, scale range, and dial size was investigated. The dials employed were scaled from 0 to 50, 0 to 100, 0 to 200, O to 400, and O to 600 units, and included graduations by tens, fives, twos, and units. Dial sizes were 2.8 in. and 1.4 in. The precision of scale reading varied with the distance on the scale allocated to each scale unit. Reading precision improved as this distance increases to about 0.05 in., a value which varied only slightly with graduntion scheme but which appeared to vary with graduation mark thickness. For more expanded scales, reading precision remained reasonably uniform at a level which depends on graduation scheme.

DISTRIBUTION: Copies of this report obtainable from CADO.

DIVISION: Aviation Medicine (19)
SECTION: General (0)

SUBJECT HEADINGS: Instrument dials - Legibility

CADO TECHNICAL INDEX

USAF Contr. No. W33-038-ac-14480