UMTRI-88-37

78868

HUMAN FACTORS AND GAUGE DESIGN: A LITERATURE REVIEW

Paul Green

TECHNICAL REPORT

UMTRI The University of Michigan
Transportation Research Institute

Technical Report Documentation Page

1. Report No.
UMTRI-88-37

2. Government Accession No.

3. Recipient's Catalog No.

4. Title and Subtitle
HUMAN FACTORS AND GAUGE DESIGN:
A LITERATURE REVIEW

5. Report Date
September, 1988

6. Performing Organization Code
389214

7. Author(s)
Paul Green

8. Performing Organization Report No.
UMTRI-88-37

9. Performing Organization Name and Address
The University of Michigan
Transportation Research Institute
2901 Baxter Road
Ann Arbor, MI 48109-2150 U.S.A.

10. Work Unit No. (TRAIS)

11. Contract or Grant No.
DRDA 86-1909-P1

12. Sponsoring Agency Name and Address
Chrysler Motors Corporation
R&D Programs Administration
12000 Chrysler Drive
Highland Park, MI 48288-1118

13. Type of Report and Period Covered
January 1, 1987 -
August 31, 1988
(Final Report)

14. Sponsoring Agency Code
2000530

15. Supplementary Notes
Supported by the Chrysler Challenge Fund

16. Abstract

This report reviews the literature on human factors/ergonomics and the design of gauges for automotive instrument panels. It covers both the general human factors literature and the literature specific to automotive applications. The conclusions of this report are as follows:

1. The best display depends upon the task to be performed. Use numeric displays for quantitative reading, pointer displays for check reading. Speedometers should be numeric displays. Engine and fuel gauges should be pointer displays.
2. The key variables for pointer displays are numerical progression and interpolation. Other factors such as scale unit length, scale orientation, marker width and length, pointer design, scale number location, etc., are secondary.

Further, this report contains numerous specific recommendations for scale mark sizes, scale numbering schemes, zone coding, and pointer design.

17. Key Words
Human factors, ergonomics, human engineering, displays, legibility instrument panels, automobiles, cars, engineering psychology

18. Distribution Statement

19. Security Classif. (of this report)
Unclassified

20. Security Classif. (of this page)
Unclassified

21. No. of Pages
146

22. Price

Reproduction of completed page authorized

PREFACE

This report describes the final task in a four-task project entitled "Recognition and Comprehension of Electronic Display Graphics." This research was funded by the Chrysler Corporation through the Chrysler Challenge Fund. The purpose of the Fund is to establish closer ties between the Chrysler Corporation and leading American universities, and to promote direct access to the advanced technologies being developed in universities. It also aims to increase interaction between the Chrysler engineering staff and university research personnel, and to increase undergraduate and graduate student awareness of the engineering opportunities available at the Chrysler Corporation.

This project is intended to provide information that designers and engineers can use to develop legible and understandable automotive displays. This particular report reviews the research on human factors/ergonomics and the design of gauges.

Other reports sponsored by this project include reviews of the literature on display legibility (task 1, 3 reports), several experiments concerned with alternative methods for evaluating legibility (task 2, 2 reports), and an experiment on the legibility of seven-segment numeric displays (task 3, 1 report).

We would like to thank Cathy Colosimo of the Chrysler Corporation for serving as the liaison for this project. Her patience and understanding were greatly appreciated. We would also like to thank Tom Dunn for his insight. Finally, we would like to thank Jim Geschke who was the initial contact person in 1984 when we approached Chrysler about this research and saw that it was included in the Challenge Fund Program.

- PREFACE -

TABLE OF CONTENTS

- TABLE OF CONTENTS -

LIST OF TABLES

- *LIST OF TABLES* -

LIST OF FIGURES

EXECUTIVE SUMMARY

Green, P. (1988). Literature Review: Human Factors and Gauge Design (UMTRI technical report 88-37). Ann Arbor, MI: The University of Michigan Transportation Research Institute.

This report concerns the final task in a four-task project entitled "Recognition and Comprehension of Electronic Display Graphics." This research was funded by the Chrysler Corporation through the Chrysler Challenge Fund. The purpose of this project was to provide information that designers and engineers can use to design displays that are legible and understandable, and, consequently, easy to use.

This report reviews the literature on human factors/ergonomics and the design of gauges for automotive instrument panels. It includes studies both from the general human factors literature and from the literature specific to automotive applications. In particular, this review addresses the following questions:

Which kinds of displays are best for various tasks?

Answer: For quantitative reading use numeric displays. Moving pointer displays are strongly preferred when the task is check reading. When used for quantitative reading, the format of the display (circular, arc, horizontal, vertical, etc.) has a minimal effect on performance. There is a trend for vertical and, to some extent, horizontal displays, to be slightly more difficult to read than circular displays. The report describes at great length several experiments that measure (in terms of errors and reading time) the benefits of each type of display.

Should speedometers be pointer or numeric displays?

Answer: Numeric speedometers are read more rapidly and accurately. This has been found to be true in laboratory studies involving slides of displays, studies conducted in driving simulators, and on-the-road studies. The basic finding has been replicated by several investigators.

Is pointer alignment beneficial for check reading?

Answer: When multiple displays are to be checked (e.g., several engine displays), aligning the pointers so they are parallel markedly reduces reading time and errors. This occurs only when the gauges are close to each other. The particular clock position of the aligned pointers (e.g., 9 vs 12 o'clock) is unimportant.

How should gauge scales be marked?

Answer: Generally, scales should be numbered in multiples of 10 (0, 10, 20,.. or 0, 100, 200,...) since people are used to dealing with the decimal system. For scales displaying values less than one, a multiplier should be used to reduce mental computation. It is not clear if speedometers should be marked in 10s (0, 10, 20,...) or 5s (5, 15, 25,...). That question needs to be examined experimentally.

The rule of thumb is that marked intervals (where the numbers appear) should be at least 1/2 inch apart, though some have argued for much larger separations. Tick marks should be at least .03 inches wide and separated by at least 1/10 inch.

How should zones on scales be coded?

Answer: Displays used for check reading should include marks (ok, normal, etc.) to indicate normal values. Color coding should also be provided. Independently, both are equally effective (as measured by the reduction in error rates in misreading displays). Other language-free marks to code displays have not been examined extensively.

How should pointers be designed?

When used for quantitative reading, pointers should come within 1/4 inch of the scale marks. When used for check reading multiple displays, much longer pointers are desired. If the scale marks are all one color, the pointer should be a contrasting color. When scale marks are multicolored, the choice of a color for the pointer is unclear.

What really matters in gauge design?

Numerical progression and interpolation have the largest effect on the time required and on the errors made in reading displays. Factors such as scale unit length and width, scale orientation, marker width and length, clutter, pointer design, scale number location, and so forth have secondary effects.

INTRODUCTION

Scope

This report reviews the literature on human factors/ergonomics and the design of gauges for automotive instrument panels. Included in this review are studies both from the general human factors literature and from the literature specific to automotive applications. The goal of this report is to provide information designers and engineers need to design instrument panel displays that will be both legible and understandable, and consequently, easy to use. This review examines the following questions:

Which kinds of displays are best for various tasks?

Should speedometers be pointer or numeric displays?

Is pointer alignment an aid in check reading?

How should gauge scales be marked?

How should zones on scales be coded?

How should pointers be designed?

What really matters in gauge design?

A chapter in this report is devoted to each of these questions. Within each chapter the question is addressed through an extensive discussion and review of previous research. Based on that research, specific recommendations aimed at answering these questions is provided.

Background

The subject of instrumentation design is a topic that has been intensively examined since World War II. Probably the best known early research on displays is the work of Fitts and Jones (1947). They interviewed 50 pilots individually, another 50 in small groups, and sent printed forms to another 524 (of which 187 were returned). The pilots were asked "Describe in detail some error which you have made in reading or interpreting an aircraft instrument, detecting a signal, or understanding instructions; or describe such an error made by another individual whom you were watching at the time." From them, reports of 270 errors were obtained, such as the following example.

- *INTRODUCTION* -

Confusing Carburetor Air Temperature and Gas Gage

> *I was copilot one night in a C-47 and the pilot asked me how much fuel we had. I said half a tank on the particular tank we were using. Thirty minutes later, he asked me again and I thought it still read half a tank. I thought something was wrong with the gage. I asked the pilot if he had the fuel selector set right and he replied 'yes'. The reason for the error was that the carburetor air temperature was directly below the gas gage and the needle was pointing straight up and down. I had mistaken it for the gas gage.* (Fitts and Jones, 1947, p. 37.)

Shown in Table 1 is a summary of the results of the Fitts and Jones study. Notice that the largest number of errors is associated with reading multipointer altimeters, obviously not a problem in cars. However, other errors, such as those due to legibility and scale interpretation can and do occur in automobiles. Those interested in the origins of research in this area should read the original report. The pilots' descriptions of errors are most interesting.

Table 1, Errors Made in Reading and Interpreting Aircraft Instruments (Fitts and Jones, 1947)

Description of Error	% Reporting
Error in interpreting multirevolution/ multipointer displays (mostly altimeters).	18
Reversal errors: Interpreting an instrument in the opposite manner to which it was intended so that subsequent actions make matters worse.	17
Signal interpretation errors: Misunderstanding hand signals, warning horns or lights, or radio range signals.	14
Legibility errors: Difficulty in reading numbers or scales.	14
Substitution errors: Mistaking one instrument for another.	13
Using an instrument that is inoperative.	9
Scale interpretation errors.	6
Errors due to illusions.	5
Forgetting errors: Failing to check an instrument before takeoff or during flight.	4

Another critical early work is Grether's 1948 paper on displays in the SAE Quarterly Transactions. That paper reviewed Fitts and Jones 1947 report, as well as several studies concerned with check reading and pointer alignment, control-display compatibility, graduation spacing on scales and errors, and so forth. It is the first publication relating to the readability of instrument panel displays to appear in the automotive literature. Other noteworthy summaries written at about that time include Kappauf's chapter on the design and use of instruments in Human Factors in Undersea Warfare (Kappauf, 1949) and the chapters on instrument dials and legibility and visual displays in Applied Experimental Psychology (Chapanis, Garner, and Morgan, 1949). References to a number of other relevant works appear in McFarland's 1955 SAE paper (McFarland, 1955).

Every major human factors document written since then has contained at least one section or chapter on display design. That includes textbooks (e.g., Bailey, 1982; Kantowitz and Sorkin, 1983; Sanders and McCormick, 1987), reference books (Salvendy, 1987; Van Cott and Kinkade, 1972; U.S. Air Force, 1987; Woodson, 1981), and design standards (U.S. Department of Defense, 1981a; 1981b). There have been several technical reports that have reviewed the literature on human factors and displays as well (e.g., Semple, Heapy, Conway, and Burnette, 1971; Heglin, 1973). The research covered in those sources is described throughout this report. Despite the wealth of material, there are no extensive reviews of displays that specifically address automotive applications, which is the focus of this review.

Considerable effort has been made to critique the studies of interest, not just to report them. Readers should bear in mind that all studies are not equal. There are many instances where the evidence from one carefully done experiment will outweigh the evidence from a half a dozen others. Some of the issues pertaining to those judgments are described in the two subsections that follow. This discussion is intended primarily for engineers and designers who have not had formal training in human factors engineering and for those who have only had a single course in the subject and limited practical experience.

The Utility of Old Data

Several studies referred to in this report were completed shortly after World War II. Those unfamiliar with research may be tempted to ignore such studies because of their age. To do so solely for that reason would be unwise. While the quality of scientific instruments has improved with time, basic principles of human behavior, and in fact, basic scientific principles in general, have not changed. As an example, Galileo's work on orbital mechanics and Newton's work on physics were done several hundred years ago but are still accepted as an accurate reflection of the world (Pool, 1988). Likewise, human factors research should be viewed the same way (Green, 1988).

What Constitutes a Good Experiment?

Judgments about the quality of human performance experiments are difficult to make. One key factor is how the data were collected. In the studies mentioned here, often a tachistoscope was used to test what people can see in a glance. In the 1940's, t-scopes were wooden boxes (about 3 feet square by 1 foot high) with a mirror that could be quickly aimed at one of several display fields. Each field had a 5 x 8 card mounted in it. Test displays were shown for 100-200 milliseconds.

Usually subjects held down a button to view a display. Both viewing time (measured in hundredths of a second) and errors were the performance measures. Readers should note that people could tradeoff between time and errors. That is, they could spend a long time viewing a display and make few errors, or, they could shorten the viewing time and make more mistakes. For good displays, the exposure duration required is short and errors are few. For poor displays, the opposite is true. While the use of t-scopes was very common in the 50's and early 60's, they are uncommon today.

A critical weakness of applying t-scope-based data to automotive design problems is that t-scope tasks ignore visual search. For many automotive displays, the problem is not just reading the display, but finding it among the collection of items on the instrument panel. Thus, test procedures that involve looking from a distance (e.g., from a road scene) to a cluster, and then measuring time, errors, eye fixations, or steering performance are more appropriate.

Another critical experiment design issue is the particular combination of test conditions to which each person responds. Human performance experiments are classified as within-subjects or between-subjects designs. In a within-subjects design each person sees all of the test displays and conditions. In a between-subject design each person sees only a subset of them. So, for example, the people in group A might see dial 1 and those in B see only dial 2. This approach is employed where the subjects are available for a limited time. Unfortunately, when significant differences are found, one cannot be sure if those differences are due to the dials or to the two groups of people. Human factors studies consistently identify individual differences as the largest source of variability. Therefore, within-subjects designs are clearly favored. In the research reported here, those differences were minimized by testing fairly homogeneous groups (e.g., young pilots with 20/20 vision). Nonetheless, there are still individual differences and, because between-subjects studies are common, the results should be reviewed with care.

A third experiment design issue is the use of statistics. Statistics serve to summarize data and help people draw intelligent conclusions about those data. While modern procedures (e.g., F-test) were developed in the 30's, they didn't see widespread use until the mid-50's. Hence, many of the studies completed before then usually employed a series of non-independent t-tests to analyze factorial experiments and draw conclusions. According to contemporary standards, Analysis of Variance or regression analysis should be used in those situations. There are cases where re-analysis by contemporary methods might lead to altering the conclusions of a study.

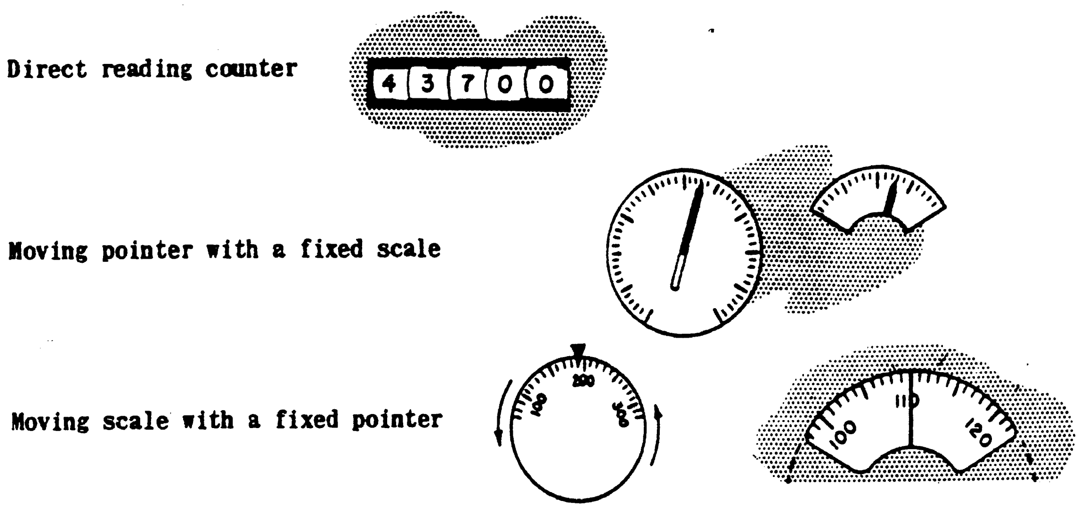
Further, readers should be cautioned that statistical differences do not necessarily mean that the differences are practically significant. Therefore, readers should look for differences that are both practically and statistically significant.

- INTRODUCTION -

WHICH KINDS OF DISPLAYS ARE BEST FOR VARIOUS TASKS?

General Recommendations for Display Selection

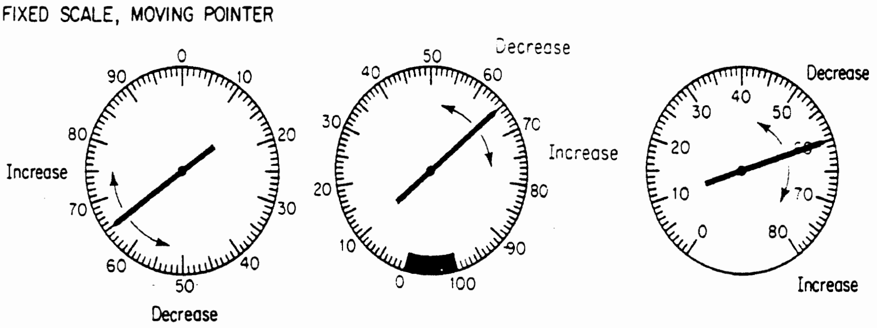
Displays are usually classified into three basic categories: counters, moving pointer displays, and moving scale displays. (See Figure 1.) A counter, now commonly referred to as a numeric or digital display, is one which shows the value of interest as a sequence of digits. A bank sign showing temperature is often of this format. In many newer cars, speedometers are numeric displays.



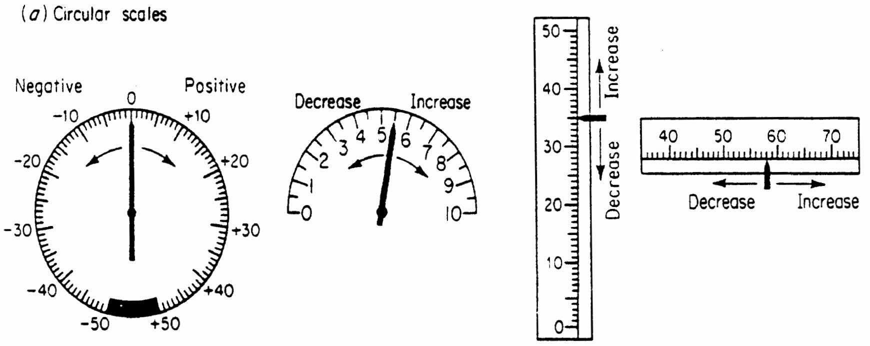
Source: Baker and Grether (1954)

Figure 1, Basic Display Types

A moving pointer display is one in which an indicator points to a scale and moves across it to identify the measured value. Most old style altimeters were of this type. In automobiles, engine gauges are usually moving pointer displays and until the 80s, most speedometers were moving pointer displays as well. Readers should note that there are many types of moving pointer displays--circular, semicircular, arc, vertical, horizontal, and so forth. (See Figure 2.)



(*a*) Circular scales



(*b*) Circular scale with positive and negative values

(*c*) Semicircular or curved scale

(*d*) Vertical scale

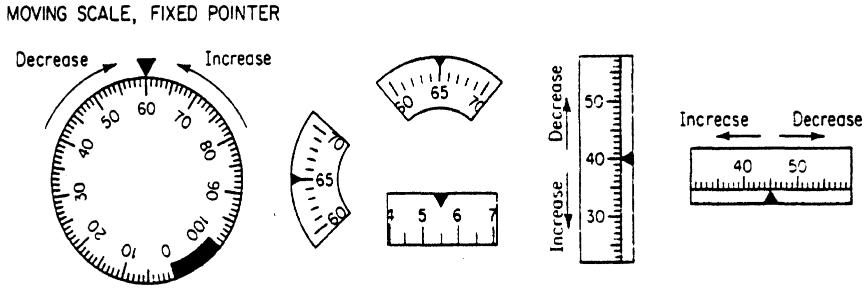
(*e*) Horizontal scale

Source: McCormick (1970)

Figure 2, Examples of Moving Pointer Displays

For a moving scale display the pointer is fixed and the scale moves behind it. Moving scale displays are not often found in automotive contexts except where the display is associated with a control used for setting, for example, to control interior temperature. At one time the Oldsmobile Toronado had a drum-type moving scale speedometer. The best known example of a moving scale display is that associated with most home thermostats. Figure 3 shows some examples of moving scale displays.

To the best of the author's knowledge, the first comprehensive set of recommendations for selection of displays appeared in Baker and Grether (1954). Those recommendations appear in Table 2. Apparently those recommendations were based on Baker and Grether's experience, having conducted research themselves, and on intimate knowledge of the literature. They do not provide specific numbers to support their recommendations though the extensive bibliography in their report suggests the studies on which their recommendations are based. The same recommendations also appear in a chapter (Grether and Baker, 1972) they wrote for a human factors textbook (Van Cott and Kinkade, 1972) that was widely used in the 70's. These recommendations are accepted practice within the profession appearing in modified form in Military Standard 1472C (U.S. Department of Defense, 1981b), Military Handbook 759 (U.S. Department of Defense, 1981a). (See Tables 3 and 4.) That information also appears in a symbolic format (Figure 4) in a variety of places (Chapanis, 1960; 1965; 1987). A sorted form of that table appears in Heglin (1973) and is reprinted in the latest edition of Sanders and McCormick, today's most commonly used human factors textbook (Sanders and McCormick, 1987)



(*f*) Circular scale

(*g*) Open-window scales

(*h*) Vertical scale

(*i*) Horizontal scale

Source: McCormick (1970)

Figure 3, Examples of Moving Scale Displays

mechanical indicators
selection of symbolic indicators

Table 2, Baker and Grether Table   Source: Baker and Grether (1954)

METHOD OF USE	MOVING POINTER	MOVING SCALE	COUNTER
1. Quantitative Reading.	Fair	Fair	Good Minimum time and error in obtaining exact numerical value.
2. Qualitative and Check Reading	Good Location of pointer easily detected. Numbers and scale need not be read. Change in position easily detected.	Poor Difficult to judge direction and magnitude of deviation without reading numbers and scale.	Poor Numbers must be read. Position changes not easily detected.
3. Setting	Good Simple and direct relation of pointer motion to motion of setting knob. Pointer position change aids monitoring.	Fair Somewhat ambiguous relation to motion of setting knob. No pointer position change to aid monitoring. Not readable during rapid setting.	Good Most accurate monitoring of numerical setting. Relation to motion of setting knob less direct than for moving pointer. Not readable during rapid setting.
4. Tracking	Good Pointer position readily monitored and controlled. Most simple relation to manual control motion.	Fair No pointer position changes to aid monitoring. Somewhat ambiguous relation to control motion. Not readable during rapid changes.	Poor No gross position changes to aid monitoring. Ambiguous relation to control motion. Not readable during rapid changes.
Comments	Requires greatest exposed and illuminated area on panel. Scale length limited unless multiple pointers are used.	Offers saving of panel space. Only small section of scale need be exposed and illuminated. Long scale possible by use of tape.	Most economical of space and illuminated area. Scale length limited only by number of counter drums.

MIL-STD-1472C
2 May 1981

Table 3, Version of Grether and Baker Table in Mil Standard

Source: U.S. Department of Defense (1981b)

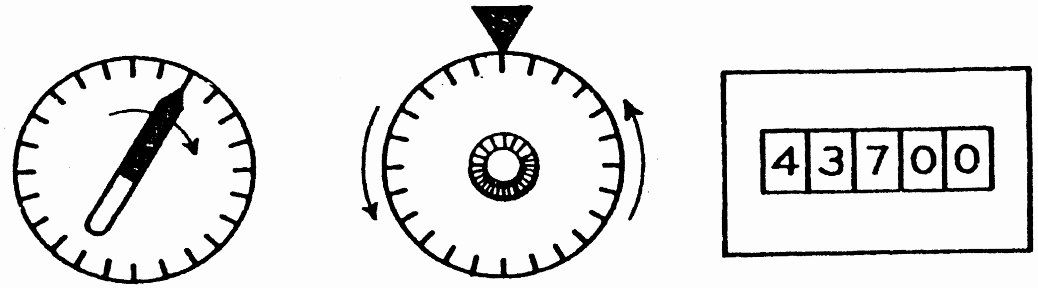
USE	SCALES		COUNTERS	PRINTERS	FLAGS
	Moving Pointer	Fixed Pointer			
QUANTITATIVE INFORMATION	FAIR May be difficult to read while pointer is in motion.	FAIR May be difficult to read while scale is in motion.	GOOD Minimum time and error for exact numerical value; however, cannot be read when changing rapidly	GOOD Minimum time and error for exact numerical value. Provides reference records.	N/A
QUALITATIVE INFORMATION	GOOD Location of pointer easy. Numbers and scale need not be read. Position change easily detected.	POOR Difficult to judge direction and magnitude of deviation without reading numbers and scale.	POOR Numbers must be read. Position changes not easily detected.	POOR Numbers must be read. Position changes not easily detected.	GOOD Easily detected. Economical of space.
SETTING	GOOD Simple and direct relation of motion of pointer to motion of setting knob. Position change aids monitoring.	FAIR Relation to motion of setting knob may be ambiguous. No pointer position change to aid monitoring. Not readable during rapid setting.	GOOD Most accurate monitoring of numerical setting. Relation to motion of setting knob less direct than for moving pointer. Not readable during rapid setting.	N/A	N/A
TRACKING	GOOD Pointer position readily controlled and monitored. Simplest relation to manual control motion.	FAIR No position changes to aid monitoring. Relation to control motion somewhat ambiguous.	POOR No gross position changes to aid monitoring	N/A	N/A
GENERAL	Requires largest exposed and illuminated area on panel. Scale length limited unless multiple pointers used.	Saves panel space. Only small section of scale need be exposed and illuminated. Use of tape allows long scale.	Most economical of space and illumination. Scale length limited only by number of counter drums.	Limited application.	Limited application

MIL-HDBK-759A
30 June 1981

Table 4, Version of Grether and Baker Table in Mil Handbook

USE	SCALAR INDICATORS		MECHANICAL COUNTERS	PICTORIAL INSTRUMENTS	PRINTERS	FLAGS
	MOVING POINTER	MOVING POINTER				
QUANTITATIVE INFORMATION	GOOD Difficult to read while pointer is in motion.	FAIR Difficult to read while scale is in motion.	GOOD Minimum time and error for exact numerical value, but difficult to read when moving.	FAIR Direction of motion/scale relations sometimes conflict, causing ambiguity in interpretation.	GOOD Minimum time and error for exact numerical value.	N/A
QUALITATIVE	GOOD Location of pointer easy. Numbers and scale need not be read. Position change easily detected.	POOR Difficult to judge direction and magnitude of deviation without reading numbers and scale.	POOR Numbers must be read. Position changes not easily detected.	GOOD Real world situation more quickly assimilated.	POOR Numbers must be read. Position changes not easily detected.	GOOD Easily detected. Economical of space.
SETTING	GOOD Simple and direct relation of motion of pointer to motion of setting knob. Position change aids monitoring.	FAIR Relation to motion of setting knob may be ambiguous. No pointer position change to aid monitoring. Not readable during rapid setting.	GOOD Most accurate monitoring of numerical setting. Relation to motion of setting knob less direct than for moving pointer. Not readable during rapid setting.	GOOD General direct control-display relationship easy to observe.	N/A	N/A
TRACKING	GOOD Pointer position readily controlled and monitored. Simplest relation to manual control motion.	FAIR No position changes to aid monitoring. Relation to control motion somewhat ambiguous.	POOR No gross position changes to aid monitoring.	GOOD Same as above.	N/A	N/A
DIFFERENCE ESTIMATION	GOOD Easy to calculate positively or negative or negatively by scanning scale.	FAIR Subject to reversal errors.	POOR Requires mental calculation.	GOOD Easy to calculate either quantitatively or qualitatively by visual inspection.	FAIR Can predict possible future pattern of pen trace.	N/A
PERMANENT RECORD	N/A	N/A	N/A	N/A	GOOD Provides hard copy.	N/A
GENERAL	Requires largest exposed and illuminated area on panel. Scale length limited unless multiple pointers used.	Saves panel space. Only small section of scale need be exposed and illuminated. Use of tape allows long scale.	Most economical of space and illumination. Scale length limited only by numbers of counter drums.			

Source: U.S. Department of Defense (1981a)



	MOVING POINTER	MOVING SCALE	COUNTER
QUANTITATIVE READING	+	+	+++
CHECK READING	+++	–	–
SETTING	+++	+	+++
TRACKING	+++	+	–

Figure 4, Chapanis Figure

Source: Chapanis (1960)

The key point these figures make is that no single display type is always best. The choice of display depends upon what the viewer will be doing with the information.

While no one has tried to directly link the specific recommendations in each cell of the matrix to specific experimental results, there is ample support in the literature for most of the recommendations.

Grether and Connell (1948)

For example, Grether and Connell (1948) conducted three experiments pertaining to display format. They examined five simulated airspeed indicators. (See Figure 5.) In the first experiment, 20 men pushed a 3-way toggle switch up if the value shown had increased from the previous one, to the right if it had not changed, and down if it had decreased. When errors occurred, they were corrected before proceeding. Each person responded 480 times to each display type (excluding practice trials). The order of displays was counterbalanced across people tested.

In the second experiment 20 men (16 who had participated in the first experiment) moved a switch to the right if a reading was the same as the previous one and to the left if it was different. All other characteristics were the same as the first experiment.

Table 5 contains the results from both experiments. The data for the right side of the dial should be ignored. They represent an experimental artifact resulting from an incompatibility between the pointer motion (down) and the switch action (up).

Table 5, Reading Times & Errors (Grether & Connell, 1948)

	Reading Type			
Display Type	Qualitative		Check	
	mean RT (secs)	Error (%)	mean RT (secs)	Error (%)
Dial-Moving Pointer				
right side	0.90	32.8	0.55	13.8
left side	0.61	6.3	0.51	9.1
Dial-Moving Scale	0.80	11.3	0.64	11.5
Vertical-Moving Ptr	0.67	10.2	0.56	12.5
Vertical-Moving Scale	0.79	12.7	0.59	13.9
Counter	0.71	9.6	0.54	10.0

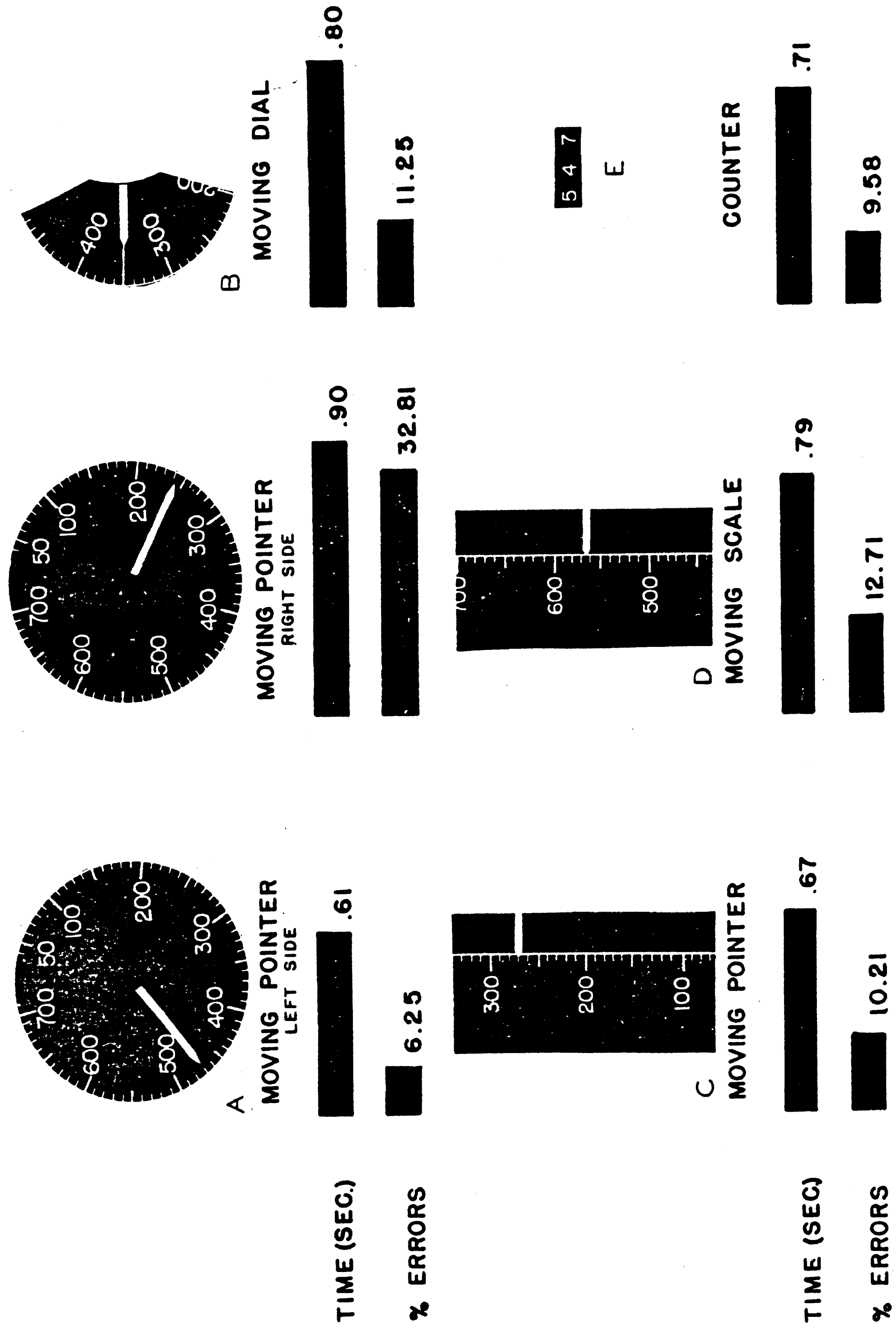


Figure 5, Displays Examined by Grether and Connell (1948)

Ignoring the right side data, the moving pointer display had the shortest response times and fewest errors for qualitative readings, though counters were almost as good. (Moving scale displays were the worst.) For check reading of single displays, there were no differences between moving pointer and counter displays and small differences between that pair and the slightly worse moving scale display.

The third experiment also involved 20 people, 13 of whom had participated in previous experiments. They moved the switch in 1 of 4 directions to indicate when a dial reading was too high. Only circular moving pointer and moving scale displays (dials) were examined.

Both the pointer position (3 o'clock, 6 o'clock, etc.) and switch motion (up, down, etc.) clearly had an effect on performance. It was best when the pointer location and switch motion were compatible. Also, with regard to display design differences, the main issue here, moving pointer displays took less time to read (.74 vs .88 seconds) and were read more accurately (21.3 vs 25.1% errors).

Sleight (1948)

Sixty college students with normal visual acuity viewed the five types of displays shown in Figure 6 in a t-scope. The circular display was 2-1/2 inches in diameter. The viewing distance was not specified, but from the drawings in the paper appears to have been between two and three feet.

Displays were exposed for .12 seconds. The student's task was to say aloud the value shown. Each student saw each dial 17 times.

From best to worst (in terms of minimizing errors) the rankings were open window/horizontal moving scale (0.5%), circular/round (10.9%), semi-circular (16.6%), horizontal (27.5%), and vertical (35.5%). The moving scale display did well because the subject was usually fixating on it. That was certainly not true for the horizontal and vertical scales where the value of interest could be out of the subject's line of sight.

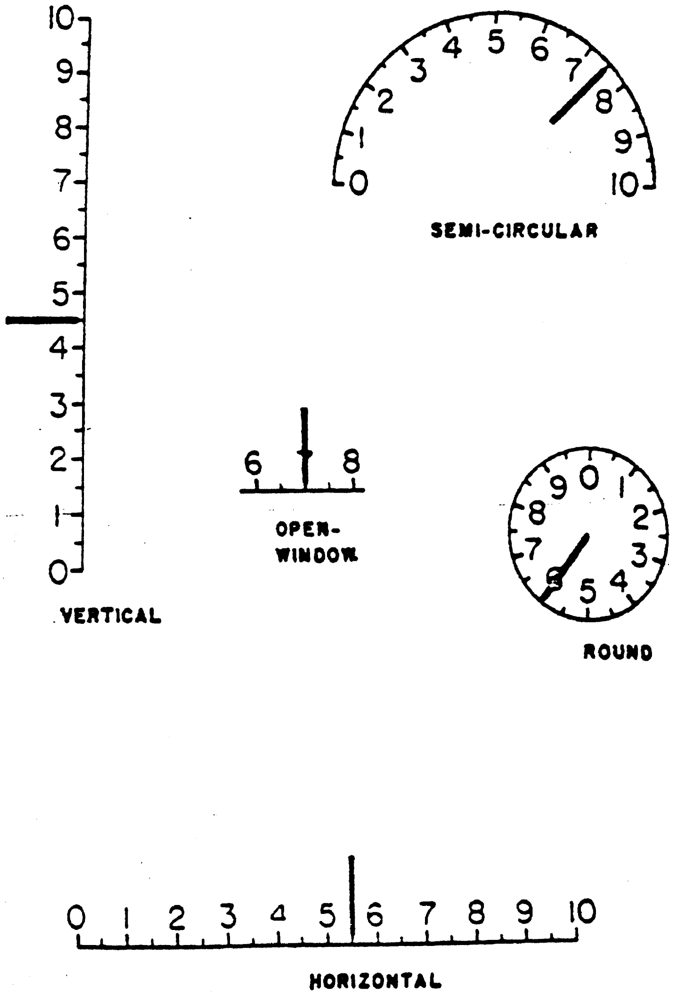


Figure 6, Displays Examined by Sleight (1948)

Connell (1950)

Connell's extension of the Grether and Connell (1948) study concerned check reading multiple instruments (such as might be found on a four-engine aircraft). Shown in Figure 7 are the moving pointer, moving scale, and counter displays she examined as they were arranged in the experiment.

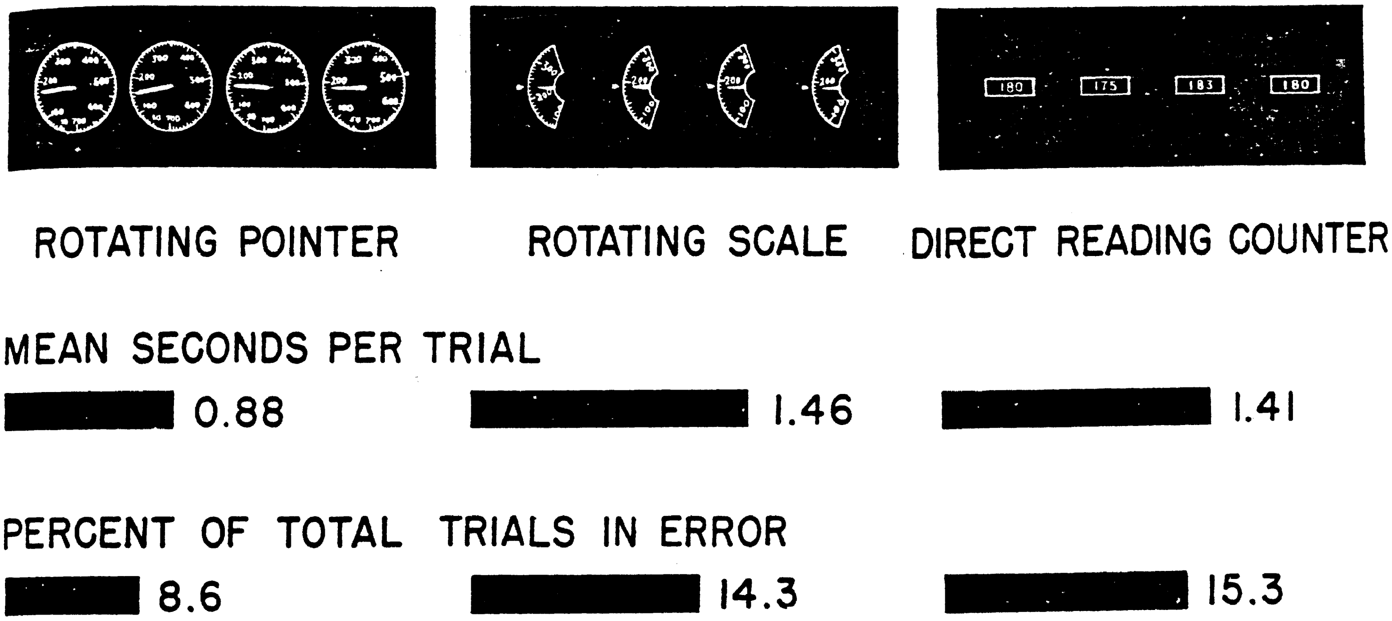


Figure 7, Displays Examined by Connell (1950)

Twenty students saw each gauge panel 15 times. They looked at each array of 4, and moved a toggle switch one way if all four gauges were the same, and another way if they weren't.

The mean time to read the moving pointer display (.88 s) was significantly less than that for the moving scale (1.46 s) and the counter (1.41 s). Error rates were also lower (8.6 vs 14.3 vs. 15.3%). Differences between the moving scale and counter were not statistically significant.

In a subsequent experiment 10 men check-read single counter displays. The experimenter said the name of a digit group aloud and then showed the participant a counter, who responded accordingly. Table 6 contains the performance data as a function of the number of digits to be checked. Both time and error rates increased linearly with the length of the digit string. The interaction of number range with display type was not examined.

Table 6, Check Reading of Digits from Connell (1950)

# of digits	Mean RT (secs)	Errors (%)
2	0.64	2.5
3	0.75	7.5
4	0.86	10.0
5	1.04	11.6
6	1.29	12.5
7	1.45	10.0

Graham (1956)

Sixty engineering students were shown a series of 1/2 second segments of film of the displays in Figure 8. After each segment they wrote down the scale reading. Each participant saw each display 10 times.

The dial tested was 5.1 inches in diameter and the linear scale 16 inches long. They were viewed at 40 inches, somewhat greater than the standard panel viewing distance.

The error rates were 31% for the horizontal scale, 35% for the circular scale, and 46% for the vertical scale. These error rates are extremely high. Only the differences between the vertical scale and the other scales were statistically significant. The circular scales had the lowest error rates because the pointer tip (where an observer must look) is on average closer to the fixation point the center of the display (where the observer is instructed to look first) than for other scales. With regard to the vertical versus horizontal comparison, the visual field tends to be wider than it is tall since people's eyes are placed side-by-side. Therefore, a pointer on a horizontal display is more likely to be within the visual field and readily seen.

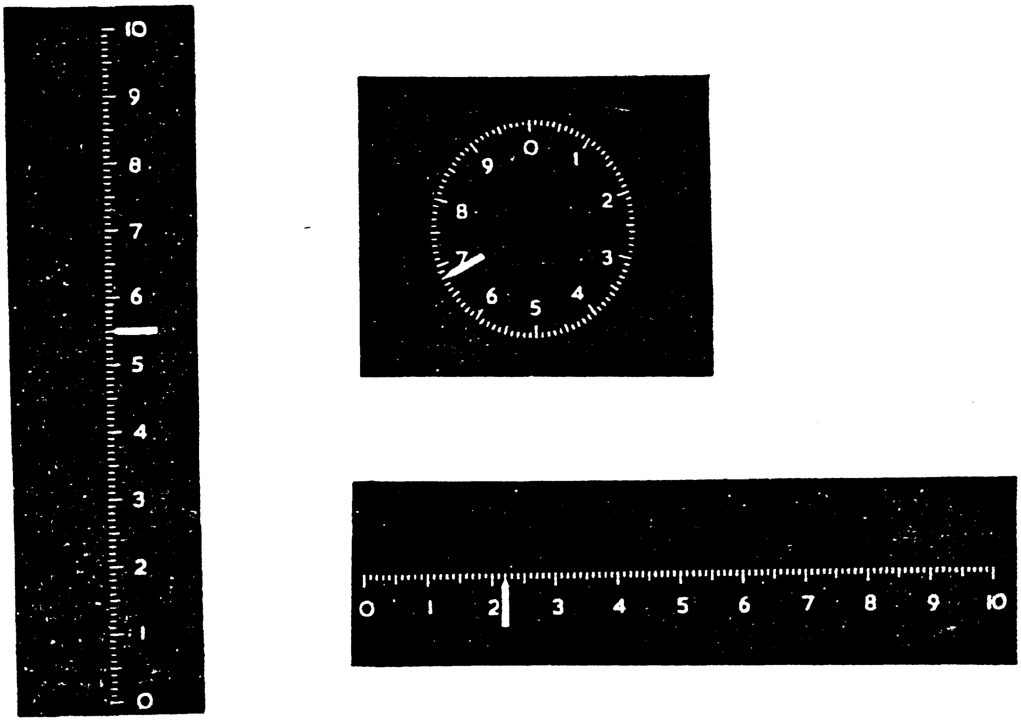


Figure 8, Displays Examined by Graham (1956)

Elkin (1959)

Elkin (1959) reported two experiments concerning the issue of display format. Vertical and circular moving scale displays and open window displays were tested using a sliding mirror tachistoscope. (See Figure 9 for examples of the scales tested.) Scales were numbered from 0 to 100 with markings every 1 or 5 units. Scales were shown for either 120, 360, 1080 milliseconds (both experiments) or until a test participant responded (first experiment only). Twelve college students with 20/20 vision or better volunteered to participate in both experiments.

In the first experiment, the participant read the display to either the nearest 5 or the nearest unit for types of markings (20 categories-nearest 5 and 100 categories-nearest unit). In the second experiment the participant said if the displayed value was high, low, or ok. Response time (from stimulus presentation to the response onset), answer time (from when the participant began to speak until they finished), and errors were recorded.

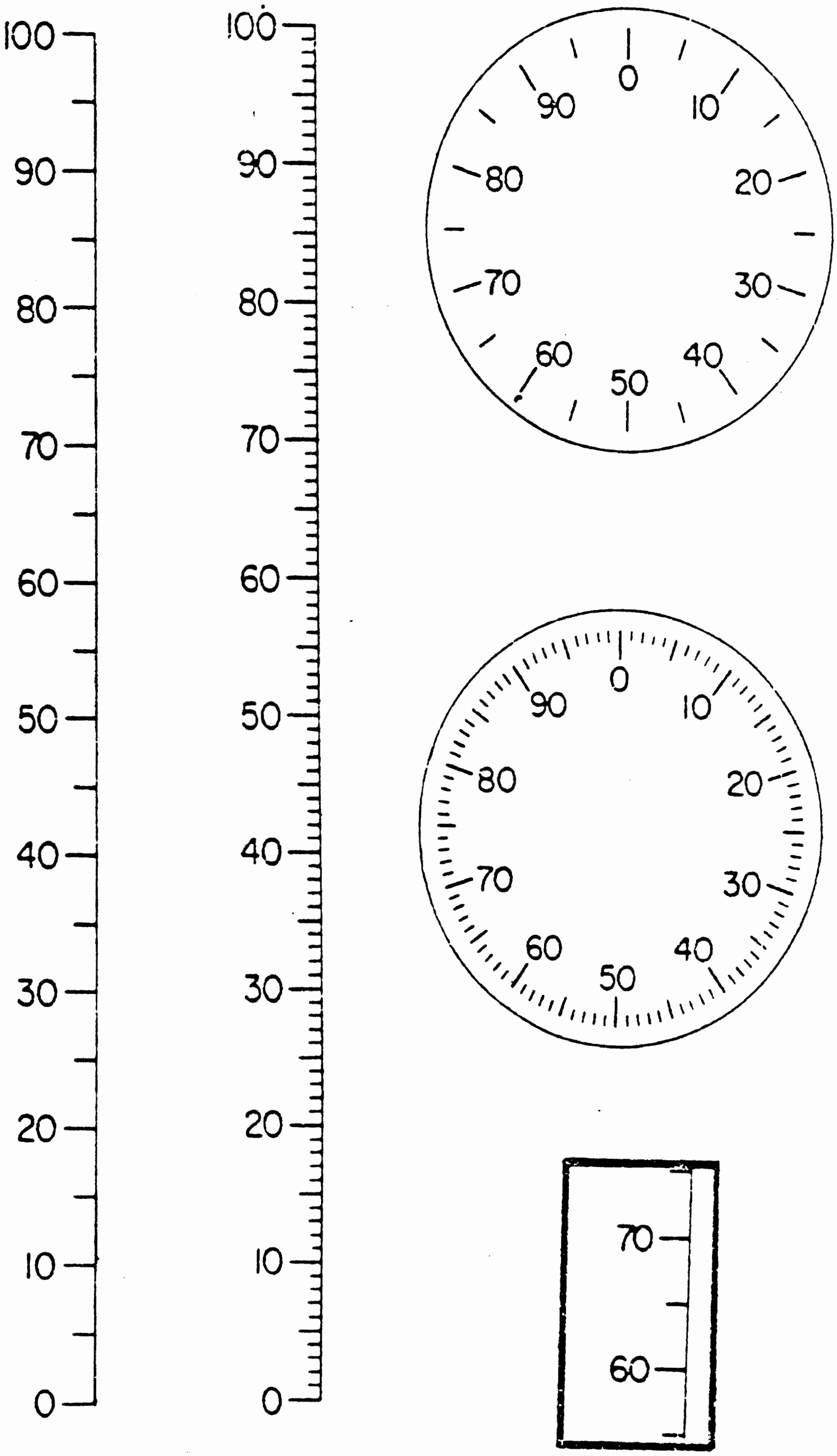


Figure 9, Scale Designs Examined by Elkin (1959)

Shown in Table 7 are the results with regard to scale type and task. The answer times have not been included because the differences tended to be less clear cut. For experiment 2, error data are not shown because very few errors were made. When making quantitative decisions people generally did better (took less time to respond and made fewer errors) with an open window display, but when the task was qualitative, a moving scale display was superior. The moving scale display was best for quantitative readings because the window format reduces the amount of search required to find the number of interest. However, when making qualitative decisions, the relative position of the pointer is used, information that is not available in an open window display.

Table 7, Interaction Between Task Type and Scale Type

Scale Type	Response Data: quantitative (experiment 1) RT (secs)	E(%)	qualitative (experiment 2) RT (secs)
Moving Scale	.77	8.1	.82
Moving Pointer (circular)	.85	14.3	.67
Moving Pointer (vertical)	.96	21.0	.74

Source: Elkin (1959)

Table 8 shows the effects of accuracy requirements on performance in experiment 1 (quantitative reading). The primary factor influencing performance was the reading precision required. Display accuracy had only a small effect on performance. In general, these relationships held across all three display types.

Table 8, Interaction Between Reading Precision and Level of Scale Markings

Display Accuracy	Reading Precision: Nearest 5 RT (sec)	E (%)	Nearest 1 RT (sec)	E (%)
Nearest 5	.80	8.3	.86	18.8
Nearest 1	.81	7.9	.97	22.9

Source: Elkin (1959)

Lincoln and Cahill (1965)

Lincoln and Cahill (1965) conducted several experiments examining similar issues. They note that in many real world situations the out-of-tolerance limits change over time. Sometimes the limits appear as marks on the glass face of the meter. At other times they are memorized by the operator.

People were tested on the moving pointer displays shown in Figure 10 and the numeric displays in Figure 11. The meters were real displays used to monitor the hydraulic system of a Polaris missile during checkout operations. The response conditions were somewhat bit strange and complex. In the pitch yaw condition, the meters were used in pairs with the left and right showing pitch and the top and bottom showing yaw. The meters were assumed to be working properly if the two meters for pitch displayed about the same values and if the two meters for yaw agreed with each other. In the roll condition, all four meters were supposedly displaying the same value.

In each condition the meters started at zero and then began to drift. People made two responses. The first, the attitude response, indicated the direction of drift. (So, for example, for Figure 10 they pressed the CCW roll button to indicate counterclockwise roll.) For the second response, the subject indicated if the meters were in or out of tolerance by pressing the associated button.

In the first experiment tolerances of +/- 2 or 4 degrees and nominal values of 8, 10, 15 and 17 degrees were examined. Sixteen electronic technicians were tested in 24 trials each. They were given unlimited time to respond. In the second experiment another group of 16 electronic technicians were tested. Displays were shown for only 5 seconds. Only tolerances of +/- 2 degrees and nominal values of 15 degrees for pitch and yaw, and 10 degrees for roll were examined. In the experiments three and four, only the numeric displays were tested. In the third experiment, 24 people not familiar with electrical measurements were tested, and in experiment 4, 10 were tested. The purpose of both experiments was to examine response artifacts.

Table 9 shows the attitude response data. The differences between display types were not statistically significant in the first experiment. However, they were in the second. (Digital displays were read more quickly.) There were no significant differences in the number of errors in either case.

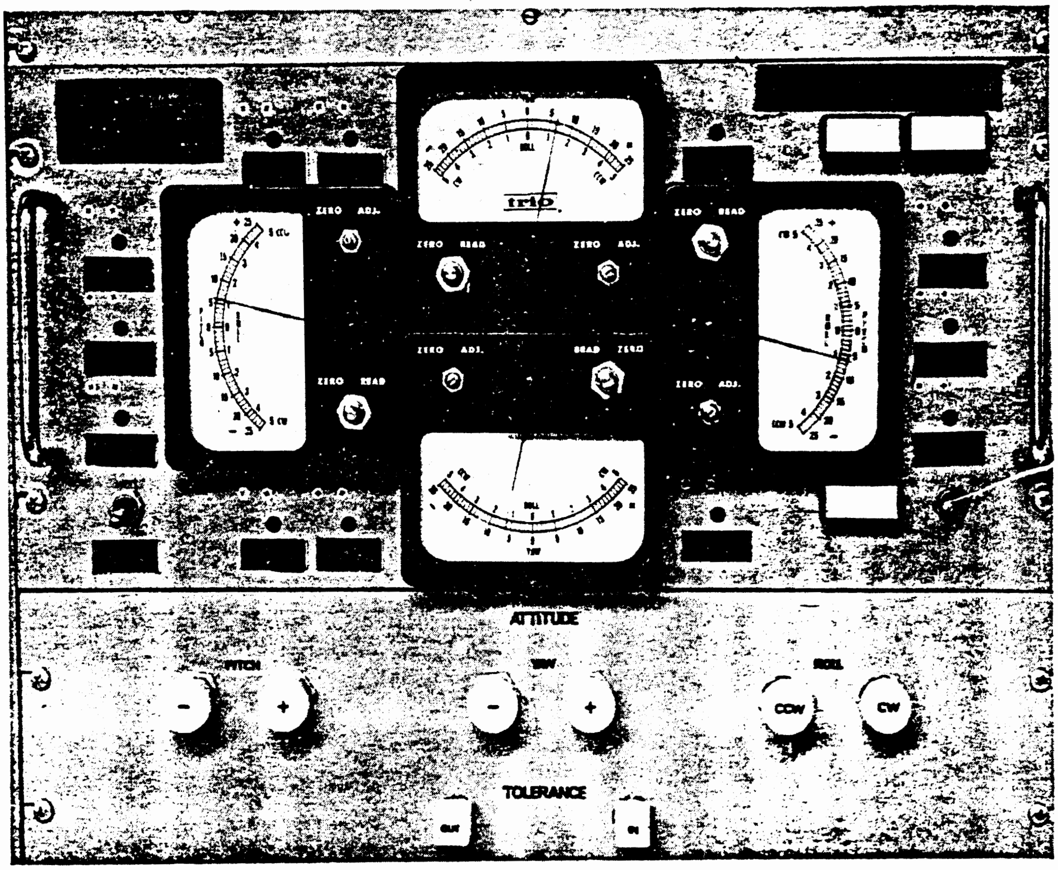


Figure 10, Meter Array Used by Lincoln and Cahill (1965)

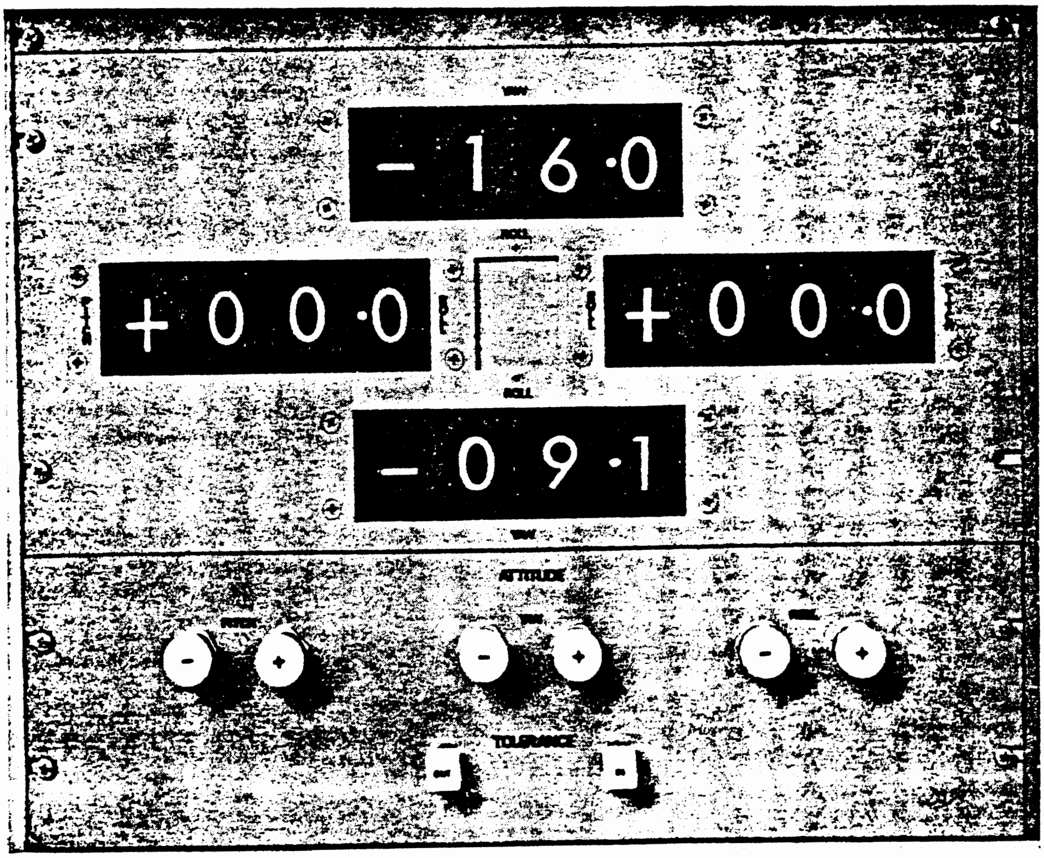


Figure 11, Numeric Displays Examined by Lincoln and Cahill (1965)

Table 9, Attitude Response Data from Lincoln and Cahill (1965)

	- Experiment 1 -		- Experiment 2 -	
Display	Mean RT (secs)	Errors (#)	Mean RT (secs)	Errors (#)
Moving Pointer	2.75	8	2.84	23
Digital	2.70	8	2.55	18

Table 10 shows the tolerance data. People responded significantly faster to the numeric display than the moving pointer display but only in the first experiment (when there was no time limit). Differences were most pronounced for small deviations. There were no statistically significant differences in terms of the number of errors, though more errors were made in responding to the numeric display than the pointer display in the second experiment.

Table 10, Tolerance Response Data from Lincoln and Cahill (1965)

	Experiment 1		Experiment 2
Display	+/-2 deg Mean RT (secs)	+/-4 deg Mean RT (secs)	+/-2 deg Mean RT (secs)
Moving Pointer	4.26	3.97	2.30
Numeric	2.62	3.10	2.10

These data are in sharp conflict with the rest of the literature. In this situation no banding was provided on the moving pointer displays to facilitate check reading (as is common practice) and the display arrangement did not permit the use of simple pointer alignment cues (discussed later). Thus, there are exceptions to the generally accepted guidelines concerning the types of tasks for which particular displays are preferred. Those exceptions occur when the displays of interest don't follow good human factors practice. Readers should also realize that the task chosen in this experiment was quite complex and confusing and these explanations could be an attempt to explain away differences that don't really exist.

Zeff (1965)

Several more recent studies have compared analog (moving pointer) and numeric clocks, a topic that generated considerable discussion when digital watches first appeared. (See Sinclair, 1971 for a review.) Zeff (1965) showed slides of clocks to 20 people. (See Figure 12.) The conventional clock was 5 inches in diameter with 1/4 inch numerals. The numeric clock had 1/2 inch numerals. The viewing distance is not given.

People wrote down the time displayed. As noted in Table 11, response time (as measured from when the slide appeared until participants began to write) for digital clocks was 1/4 that for conventional analog clocks. This difference was statistically significant. Differences due to hour format (12 vs 24 hour) were small, but present.

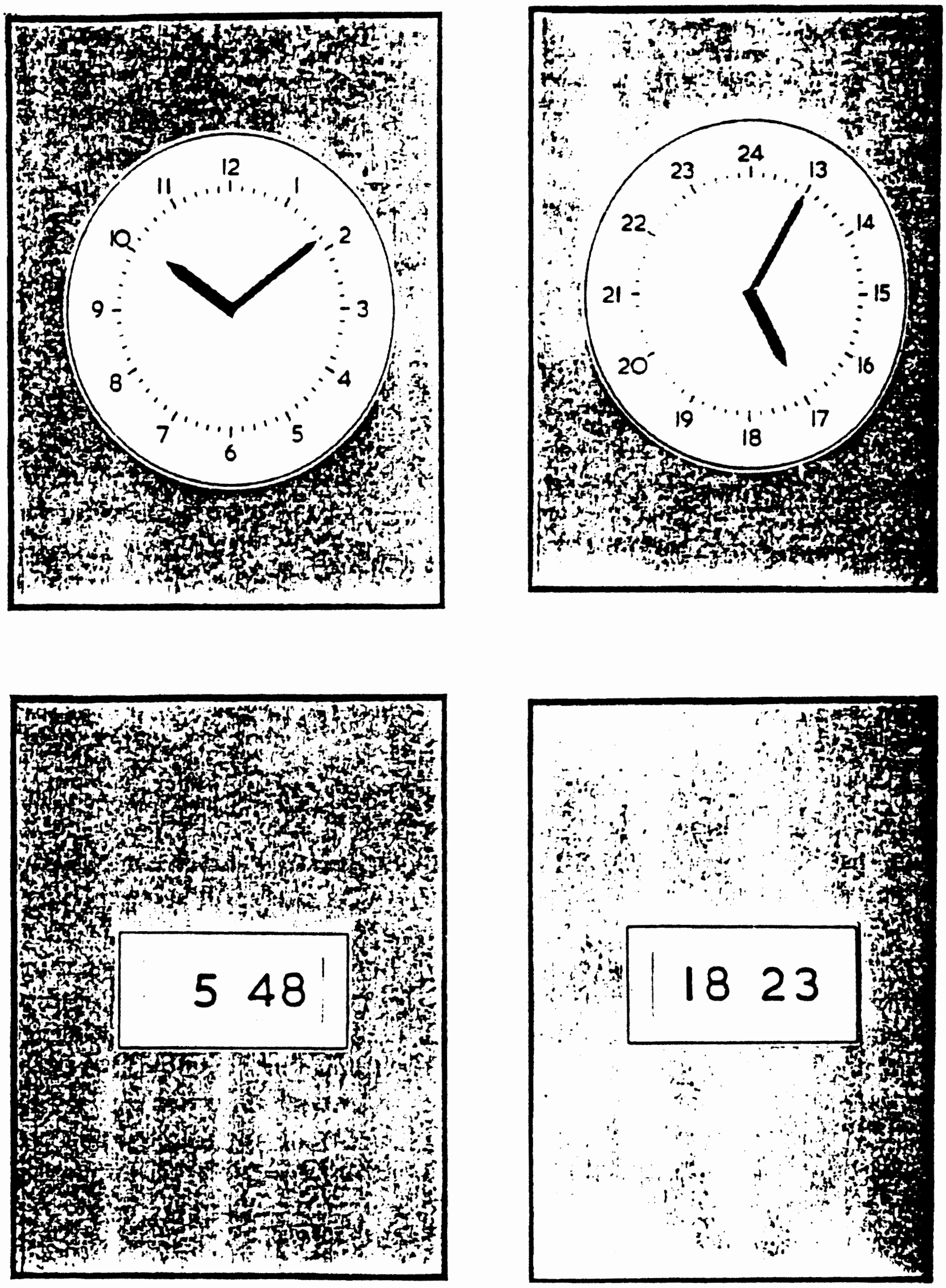


Figure 12, Clock Displays Examined by Zeff (1965)

Table 11, Mean Response Times for Clocks (seconds)

Hour Format	-- Clock Type -- Digital	Analog
0-12 hours	0.93	3.37
13-24	0.95	3.71
Mean	0.94	3.54

Source: Zeff (1965)

Cohen (1971a, 1971b)

Another study supporting the general conclusions provided earlier in this section is Cohen (1971a, 1971b). Her dissertation examined an interesting question: How well is information from a display remembered as a function of the decision to be made and the nature of the interruptions. A common example of this occurs when one looks up a telephone number and, just before dialing, is asked a question. (What time is it?) Even though the interruption is brief, people are likely to forget the number they just looked up. In an automotive context, problems take the form of forgetting information displayed on the instrument panel.

Cohen conducted four experiments concerning the relationship between the display type and short-term memory for the information shown. In the first experiment, reading errors for various exposure durations for moving pointer and moving scale displays were obtained to identify appropriate test conditions for later work.

In the second experiment, moving pointer, moving scale, and digital displays were shown to 24 students with 20/20 vision. Sample displays are show in Figure 13. Slides of displays were shown for 1.5 seconds. (In pilot studies that exposure duration lead to reading error rates of less than 5%.) Subsequently, people were shown a letter for .5 seconds, and asked to recite the alphabet backwards from that letter. Five, 10, or 20 seconds after the display was shown, participants were cued to recall the value presented. While this task may seem strange to those unfamiliar with psychological research, it is an effective and commonly accepted method that forces a person to focus on verbal information.

Shown in Figure 14 are the results. The recall of quantitative information is generally more accurate for digital displays than for other types (over time) when people are given a verbally interfering task. This may be because the digital display was easier to remember, or it could be because that display took less time to read, leaving more time to rehearse the value displayed.

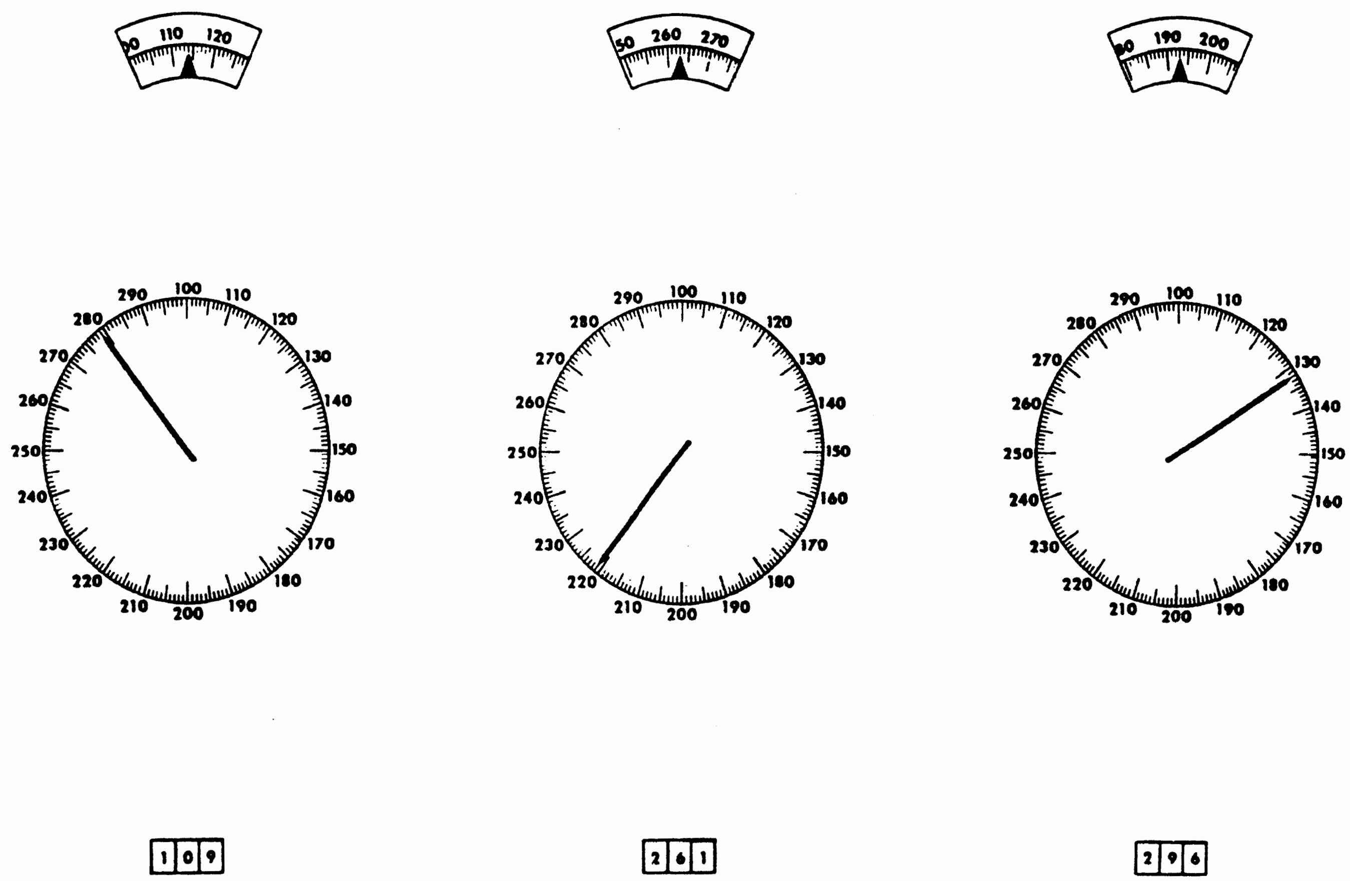
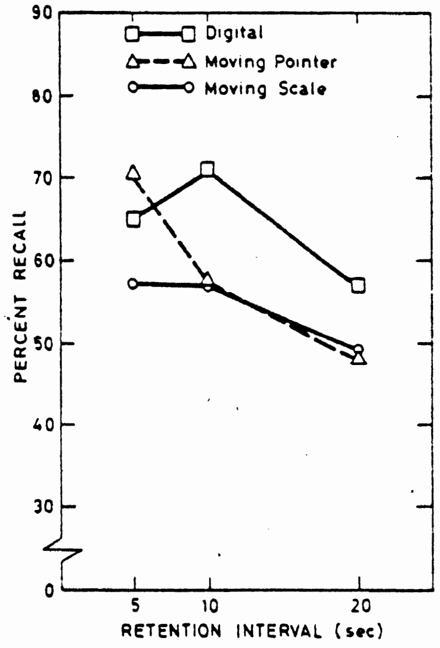


Figure 13, Displays Examined by Cohen (1971a, b)



Source: Cohen (1971b)

Recall performance at various retention intervals for readings from digital count or moving pointer and moving scale displays.

Figure 14, Results from 2nd Cohen experiment

While the moving scale and moving pointer displays had equal error rates, the size of the errors for moving pointer displays tended to be much less. In practical situations, this can be an important difference. Apparently, the pointer location for the moving pointer display provided cues that were useful to viewers.

To explore the effect of the type of interfering task on recall, another experiment was conducted involving 36 male volunteers with 20/20 vision. They were shown a test display for 1.5 seconds and then a letter for 2 seconds. In the verbal condition, they read the test display aloud (to identify reading errors) and then recited the alphabet backwards from the letter shown, as before. In the "spatial" condition, they used the 8 compass directions to identify the path required to trace out a character. For the F in Figure 15 the path is N, E, S, W, S, etc. In driving, this is like being asked to give or recall directions.

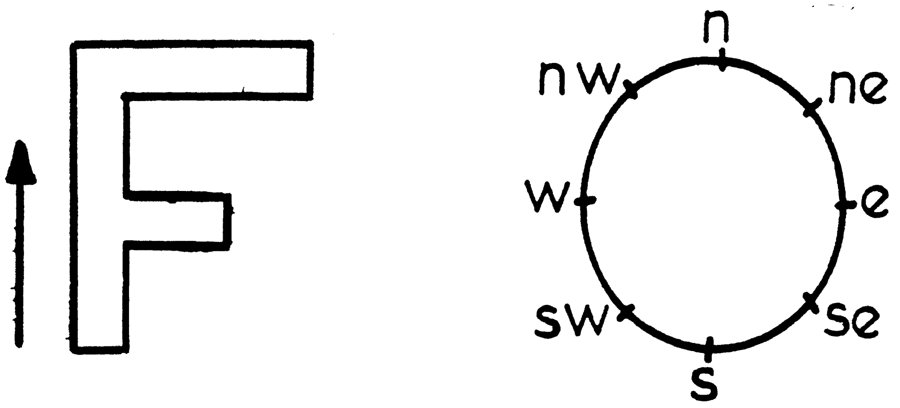


Figure 15, Sample Outline Letter from Cohen's Study

Figure 16 shows the results. Notice that differences between displays are more pronounced for displays in the spatial task with that task interfering with recall of the pointer position.

The final experiment, involving 12 male volunteers, was similar to the previous one in design except that all testing was within-subjects and the values shown on the test displays were not read aloud. The general pattern of results was similar to the previous experiment. Again, not only did the relative number of errors for each type of display change, but so did the pattern. Large errors were much more likely for the moving pointer display when the task involved spatial interference.

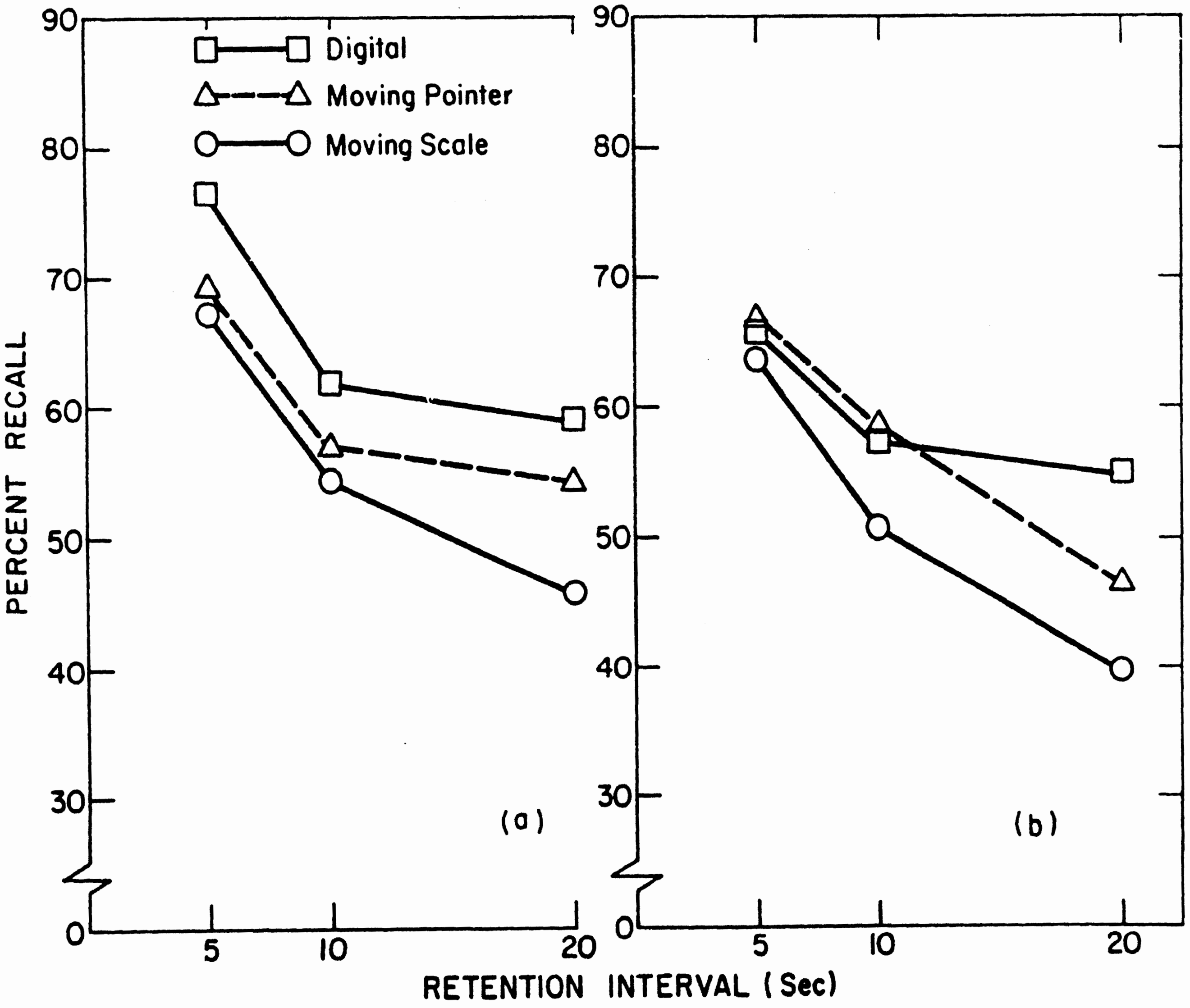


Figure 15, Error Data from Cohen's Third Experiment

Source: Cohen (1971a)

So, the critical issues are purpose the for which information is to be used (classically examined), and here, retention duration and nature of other ongoing and interpolated activities. In general, verbal activities interfered with information stored numerically (e.g., numeric displays) and spatial activities interfered with information stored spatially (moving pointer displays). For each interpolated activity, the retention interval had no effect on which information format was most likely to be remembered (though it did affect both recall probability and the size of error of recall).

Van Nes (1972)

Van Nes (1972) showed 20 people pairs of time displays and asked them to report how much later the right clock was than the left. (Figure 17 shows them at about 1/2 their actual size.) They could report the result in either minutes or hours and minutes. Half of those tested were asked to correct their mistakes.

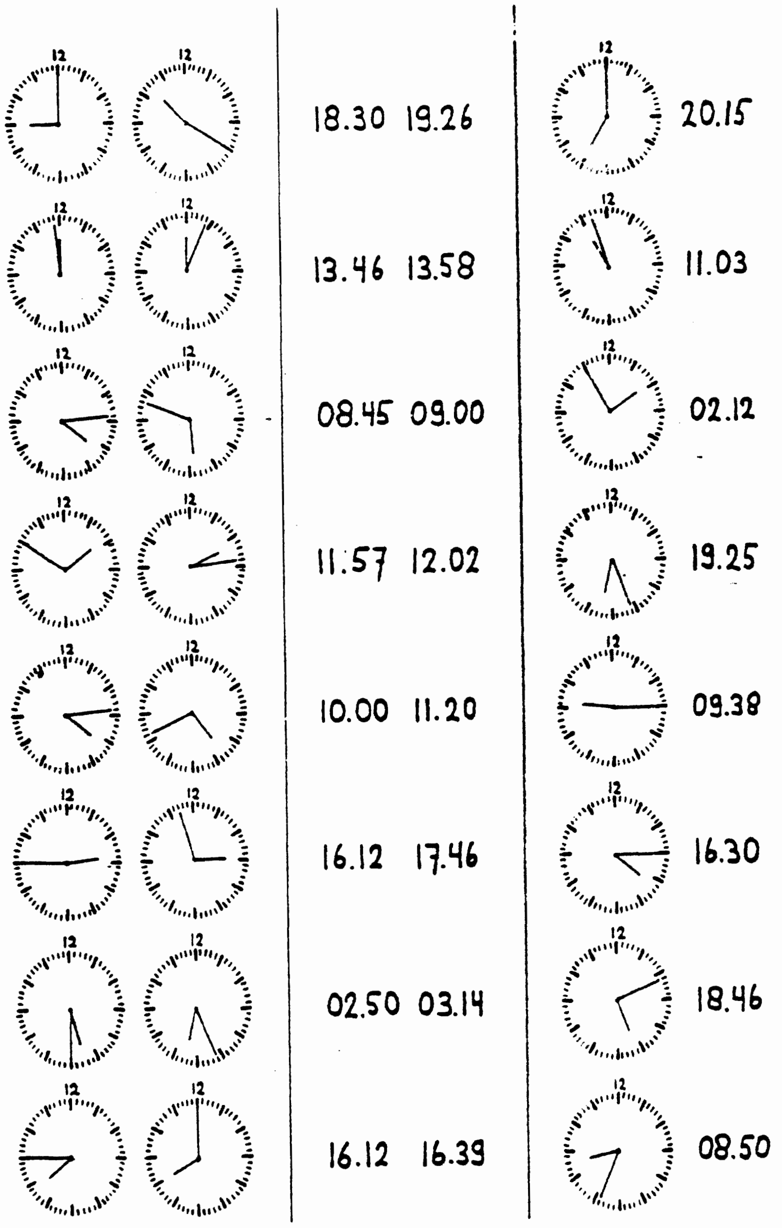
As noted in Table 12, people did best with the digital-digital pair, worse with the incompatible analog-digital pair, and worst when both clocks were analog. As a pair these studies reinforce the need to consider the task for which a display is used when choosing a display.

Table 12. Clock Reading Times from Van Nes (1972)

	Read Once Group		Read & Correct (if needed)	
Clock pair	Time (secs)	Errors (%)	Time (secs)	Errors (%)
Analog-analog	82	20	118	15
Digital-digital	46	10	51	5
Analog-digital	79	12.5	115	22.5

Summary

The research on display format supports the conclusions shown earlier in Tables 2-4 and Figure 4. The display that is "best" depends upon how it is used and to some degree, how well the particular display design follows accepted human factors practice. If the viewer is reading a display to determine an exact value, then the display should present the number required. In those cases, a numeric display should be used. If a person is checking to see if a display or group of displays are within certain bounds, then a pointer-type display is best. In general, there are few instances where a moving scale display is preferred. Additional discussion of these data appears in the section on pointer alignment and in the final summary section.



Analogue-analogue time-display pairs.

Digital-digital time-display pairs.

Analogue-digital time-display pairs.

Figure 17, Clock Displays Examined by Van Nes (1972)

With regard to the shape of moving pointer displays, circular scales tend to be better than horizontal scales which are in turn better than vertical scales. These differences are due to how far from a fixation point a pointer tip is likely to be and to the shape of the visual field. It is not clear if these differences occur in nonlaboratory tasks where visual search to find the display is required. That question should be examined experimentally.

SHOULD SPEEDOMETERS BE ANALOG OR NUMERIC DISPLAYS?

The speedometer is often thought of as the most important display on the instrument panel. Its format has been studied extensively. Some have argued that since drivers primarily want to know if they are over the speed limit, the speedometer should be a moving pointer display since that format is preferred for check reading tasks. Others argue that drivers want to know approximately how fast they are going and, consequently, the display should be numeric since the human factors literature recommends that type of display for quantitative tasks. Several studies in the literature address speedometer format directly.

Oho (1979)

Oho (1979) describes three experiments concerned with speedometers. The description of both the methods and the results is incomplete. In the first experiment ("sensory response") a small car (Isuzu) was fitted with an additional seven-segment LED display. The digits were 6.5 mm high, nearly equal to the digits on the circular moving pointer speedometer that was also provided. While a variety of problems with the initial version of the numeric display are described, no performance data are presented.

In the second experiment 20 drivers with 20/20 acuity or better participated. Speedometer-like displays (moving pointer and numeric) were viewed through a shutter. The exposure duration was .2 seconds for the moving pointer display and either .2, .5, or 1.0 second for the numeric display. The visual angle for both displays varied between about 20 and 120 minutes of arc (about .17 to .98 inches high). Their task was to look ahead at a light. When it was illuminated, they were to look down at the cluster and read the speedometer.

Percent correct for the moving pointer displays close to 100% for all character sizes. For numeric display, the percent correct was almost equal to that level for the 1.0 and 0.5 second exposure durations. Statistical tests of the differences are not provided. For the .2 second duration, percent correct varied between 90% at 80 minutes of arc and 70% at 20 minutes. The data suggest that numeric displays should be between 80 and 120 minutes of arc. (The curve in the best copy available is fuzzy, and if reproduced here, would not be readable.)

Apparently the illumination level was also varied in this experiment or experiment series for the numeric display. Details concerning those results are contained in the last of the three reports in this project concerning legibility of displays. Similarly, the third experiment, which concerned update rates and other parameters, is described in that report as well.

Galer, Baines, and Simmonds (1980)

This is the classic study of speedometer design. (See also Anonymous, 1981; Baines, Spicer, Galer, and Simmonds, 1981; Simmonds, Galer, and Baines, 1981.) Three experiments were conducted. In the first 75 drivers were shown slides of five instrument cluster designs. (See Figure 18.) They included two similar electro-mechanical designs, an electronic equivalent, an electronic curvilinear display (hockey stick design), and an electronic numeric/digital display. Details concerning the digit size, luminance, color, etc. are not provided.

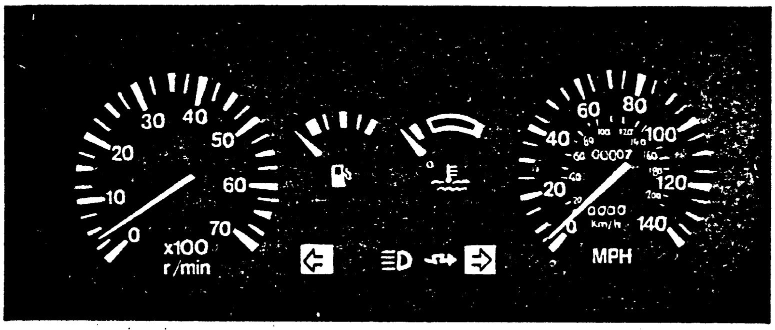
Drivers were shown 5 practice slides of each display, and then 15 test slides with each slide presented for 450 milliseconds. Two sets of slides were used. In the first only the speedometer appeared. In the second, the entire cluster was shown. In each case the driver wrote down the speed shown and noted if it was above the speed limit.

As shown in Figure 19, there were fewer errors made (2%) in identifying the speed shown on the numeric display than others. Also, the numeric display was check read at least, if not more accurately than the alternatives. The numeric display was considered easiest to read (70%) and easiest to check against the speed limits (47%). In terms of preferences for their own car 45% selected the numeric display, 25% selected the revised electro-mechanical module, and the remaining 30% selected other designs. The data indicated a strong dislike for the curvilinear display. Unfortunately, none of the documents mentioned provided any suggestions of the statistical significance of these results.

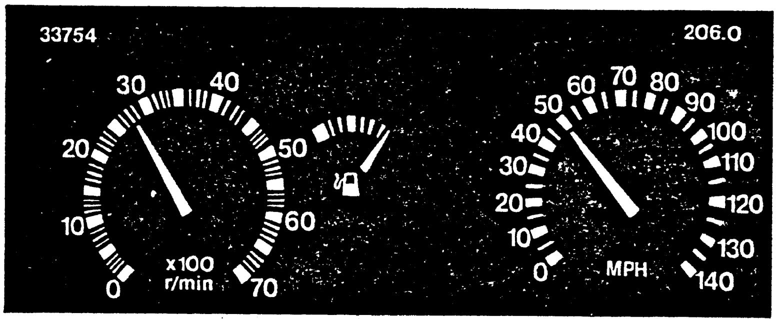
In a second experiment 100 drivers operated a driving simulator. The simulator computer also controlled the cluster displays. Except for some color changes and the exclusion of the revised electro-mechanical module, the displays tested were similar to those in the previous experiment.

In this task participants "drove" at a constant speed and performed tasks similar to that of the previous experiment, though response times were also recorded. It is not clear how drivers were cued to respond.

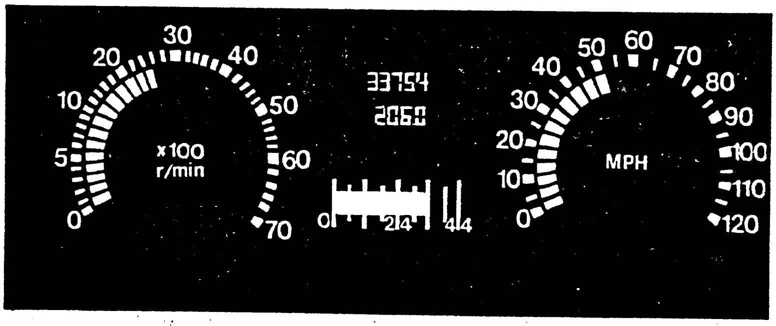
For the numeric speedometer, 97% of the readings were within 2 mph of the correct speed. (See Figure 20.) Other displays were not read as accurately. The average response time was 1.19 seconds for that display versus 1.65 for others, almost a 1/2 second difference. Statistical analysis of those data or discussion of the steering error data are not provided.



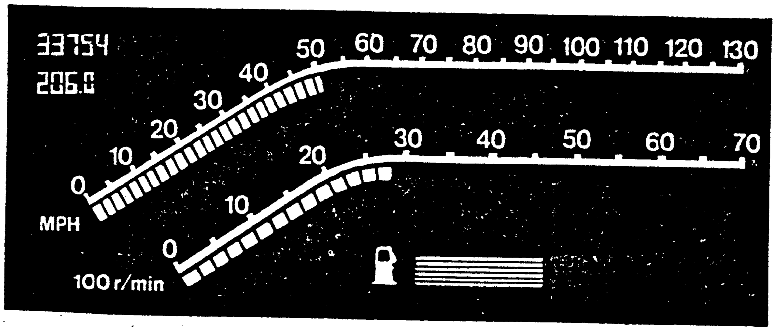
Electronic-mechanical R-module



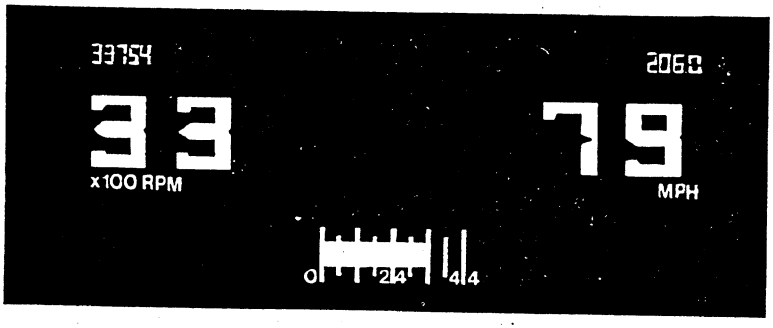
Revised electro-mechanical



Electronic dial display



Electronic curvilinear display



Electronic digital display

Figure 18, Clusters Examined by Galer et al. (1980)

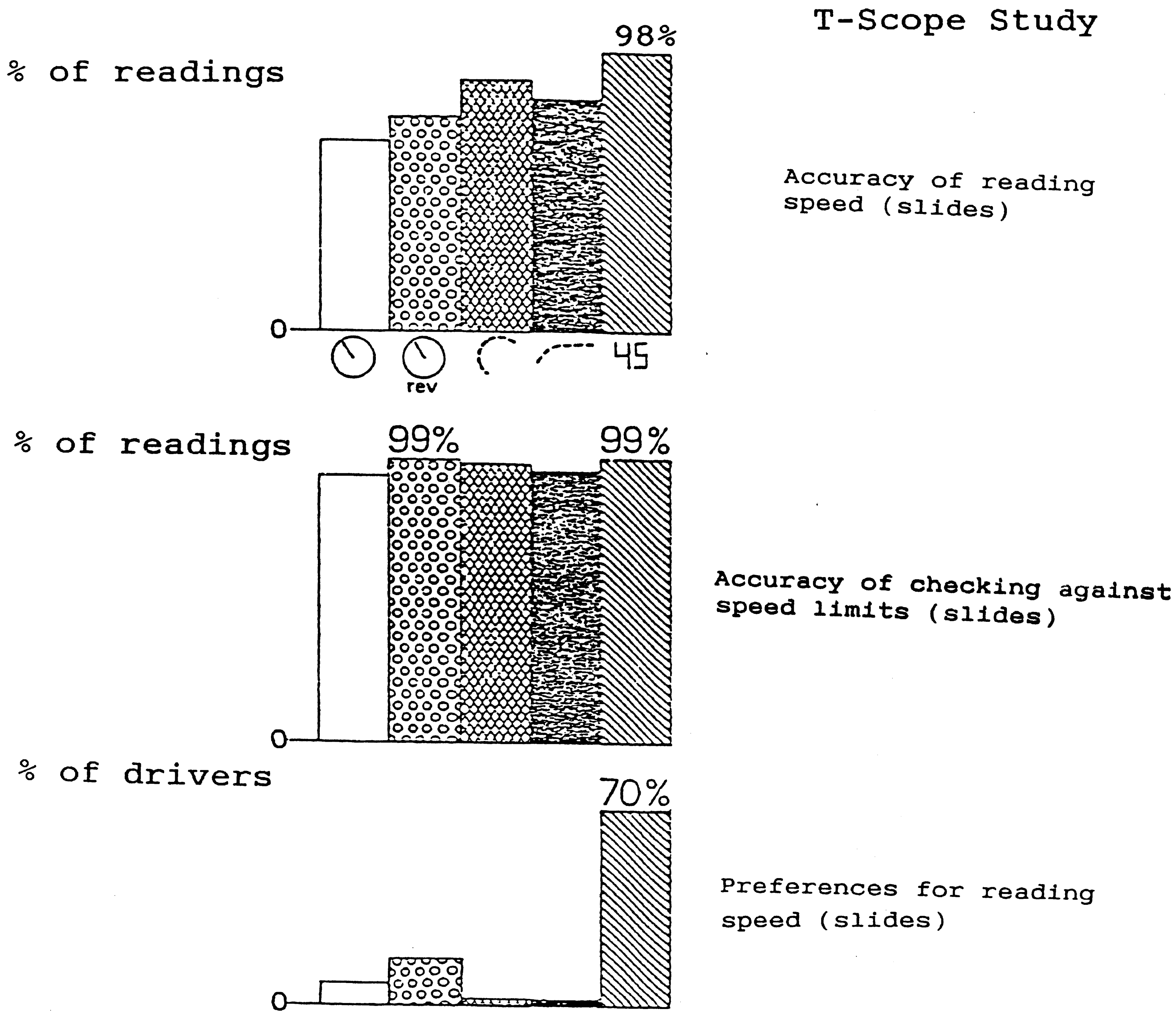


Figure 19, Results from Tachistoscopic Experiment

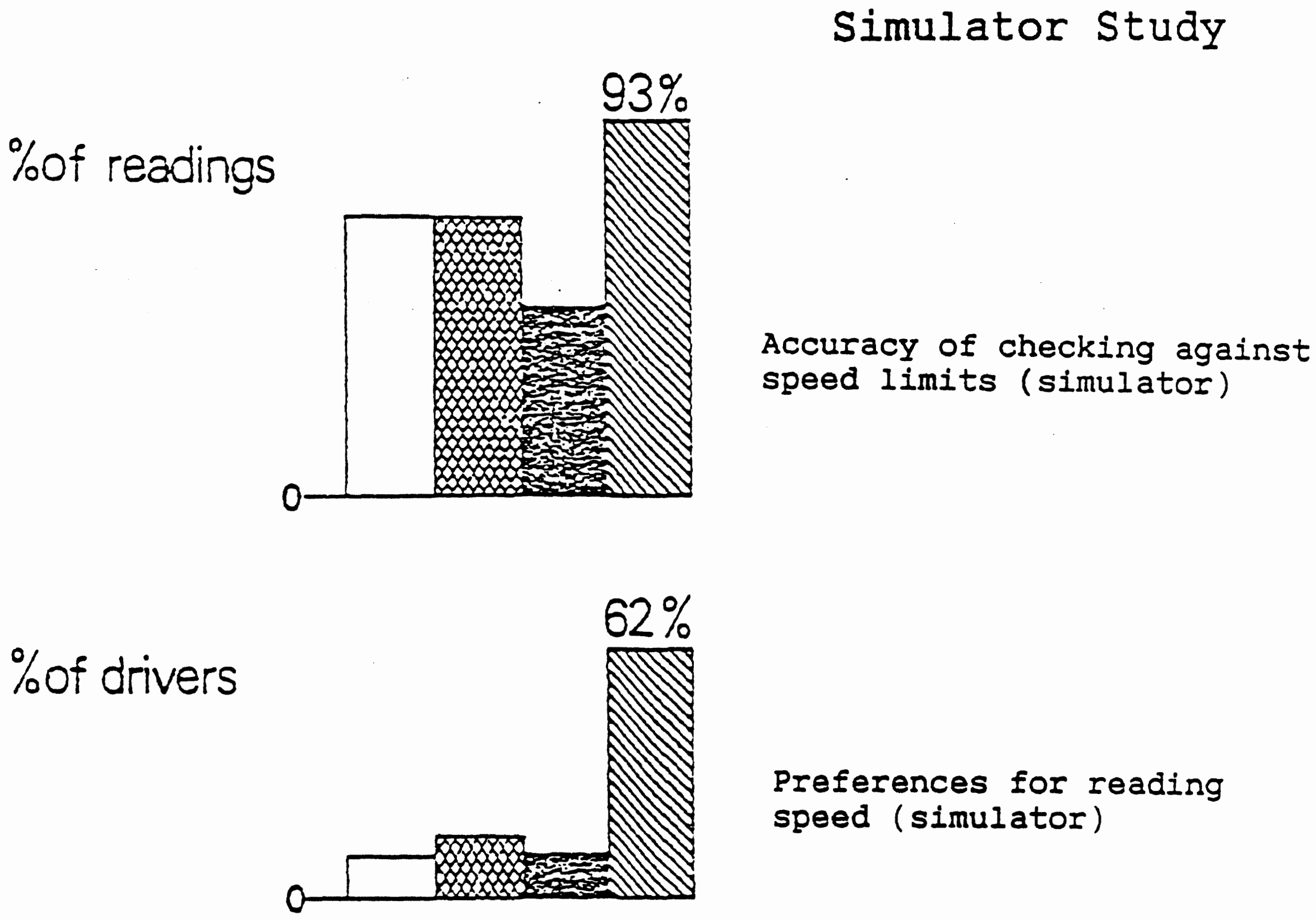


Figure 20, Results from Simulator Study

The numeric display was also preferred by drivers (67% said it was easy or very easy to drive to a target speed versus 53% for the dial, 46% for the electro-mechanical module and 43% for the curvilinear).

Those arguing against numeric displays say that changing of the numbers may be distracting (the flicker problem). About 1/5 drivers did not find any of the displays to be distracting. When problems were reported, they were most common for numeric displays. Galer, et al (1980) argue this suggests distraction is not a problem, an argument the author of this report considers to be weak.

Finally, when asked for their preference, 39% of the drivers would choose a numeric speedometer for their car. This was twice the preference of the second ranked choice.

In the third experiment, three groups of 75 drivers drove two cars. One was fitted with conventional instrumentation, the other with one of three displays: LCD implementation of the circular and curvilinear speedometers, and a simulated numeric display. The route consisted of a mixture of city, expressway, and rural driving. Tests were conducted both during the day and at night.

There were few errors made in check reading speedometers and differences between designs were not reported. However, there were differences in how accurately the speed was identified. For the daylight conditions, digital displays were more reliably read (almost 100% accurate) than conventional instruments (90% accurate). The difference was statistically significant. In the nighttime tests the electronic circular and numeric displays were read more reliably than the conventional display. Overall, the numeric speedometer was read more reliably than the electronic circular or curvilinear displays. With regard to preferences, numeric speedometers were consistently preferred over conventional instruments both for day and night driving.

Hence this research strongly supports the use of a numeric speedometer. Three studies involving about 400 people were conducted: one using slides of clusters, a second involving a driving simulator, and a third on the road. In all cases numeric speedometers were either read more rapidly or accurately than other designs, and drivers preferred the numeric displays.

Why is this so? Based on the general human factors literature, one would think that since the most common task is determining if one is speeding, a task akin to check reading, that a moving pointer display should be best. There are several reasons why numeric displays do so well. First, the numeric speedometers are unique in their design in that they are the only large numeric display provided. Since the speedometer location varies from vehicle to vehicle, this distinctiveness helps the driver find it. On the other hand, a pointer-type speedometer will at times resemble the tachometer (if provided). Second, check reading a pointer display depends upon certain conditions which are not true for the speedometer. (Pointer alignment and check reading are covered at length later in this report.) The location of the speedometer well inside the periphery and the attentional demands of driving make it impossible for the driver to gain information from the speedometer without looking at it. Also, the location of critical speeds (55 mph) varies from speedometer to speedometer and they are sometimes not located at the four cardinal positions (3, 6, 9, and 12 o'clock). Finally, the digits on numeric speedometers are big and easy to read. In the words of one of the subjects in this research, "This is the first time in years I could read the numbers on the speedometer." (Simmonds, 1981).

Ishii (1980)

This experiment compared an unknown analog display with a 7-segment numeric vacuum florescent display (VFD), both of which displayed speed in kph. The numbers were 21.5 mm (.85 in) high. Details concerning the luminance, color, etc. are not provided.

Six people drove an unknown vehicle fitted with either display. The test course included highway, suburb, and city driving. Details of the route are not provided. Data were collected for three weather conditions (clear, cloudy, rain), though not all drivers participated in all test conditions. An eye camera was used to videotape (at 30 frames/second) where the driver looked.

Shown in Figure 21 are the reading times for both displays averaged across weather categories and road types. The mean time for the numeric display was 70 milliseconds less than that for the analog display. A statistical analysis of the data (e.g., ANOVA) was not provided.

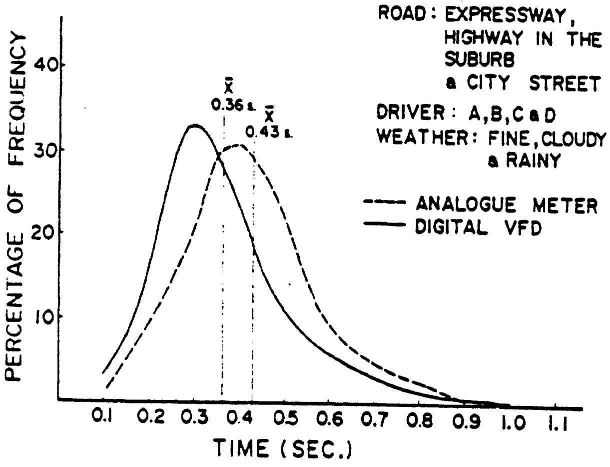


Figure 21, Data from Ishii (1980)

According to their data, drivers took more time to read the speedometer on less congested road (expressway: 460 ms (analog) vs. 410 ms (numeric), suburb: 410 vs. 340, city: 400 ms vs. 350 ms) but there was almost no change in the difference between the two display types. As one would expect, there appear to be interactions between drivers and display types. (Again, the necessary statistical analysis was not provided.)

Hence, these data all support the use of numeric displays for speedometers, reinforcing the Galer et al. data. However, there is a great deal of information about the displays, test procedure, and considerable analysis that should have been provided but weren't.

Green (undated)

In the early 1980s the author conducted a number of unpublished experiments comparing the ease of reading of various types of speedometers. Engineering students participated as a course requirement. Most had 20/20 or corrected 20/20 vision and were 19-21 years old.

Slides of instrument panels where displayed on a screen at the front of the room under computer control. Participants, tested in groups of 16, sat in two rows, roughly 8 and 16 feet from the screen respectively. Images were sized such that for those in the front row they occupied the visual angle in production vehicles. Those in the back row who saw the images at half the normal visual angle and served as the "simulated elderly" (those with degraded vision).

There were two basic tasks. In one, participants were to press one of two buttons to indicate if the speed shown was over the speed limit (55 mph). In the second task participants indicated the speed shown by pressing one of four buttons (50, 55, 60, 65 mph). Three exposure durations were used (500, 1000, or 2000 ms) with the duration selected varying from semester to semester.

Speedometers examined included the 68 Chevrolet Impala (odd scale, big numbers), 1975 Dodge Dart (rectangular speedometer), early 80's Volvo (red 55), 1981 Peugeot 505 (similar speedometer and tachometer), 1981 Renault R5 (excess scale marks), 1981 Renault R18 (speedometer marked in 20's), and a 1981 Ford Thunderbird (numeric).

Slides were shown in blocks of 16 or 48 with subjects receiving at least 4 practice blocks with slides showing speeds as words before testing began. Depending on the test condition, speedometer types could be mixed within test blocks.

Because funding was not available, a statistical analysis of these data was never carried out. Differences between designs were common and practically significant. In general, response times to the numeric speedometers were much less than those for any of the analog alternatives, though the error rates were comparable. With regard to the analog displays, the Impala speedometer often did well because it had very large numbers. Often the Volvo cluster, supposedly "ergonomically designed," did rather poorly. All of the marks on the speedometer were white on black, except for the 55 mph digits which were red and of very low luminance contrast. This attempt to highlight 55 made the speedometer more difficult to read, not less.

Armour (1984)

In a recent study by Armour, both the format and location of speed displays were examined. In work carried out at in England 22 police officers familiar with Head Up Displays (HUD) and an unknown number of members of the driving public served as participants. The police officers drove a specially fitted car around two fixed routes consisting of airfield runways and perimeter roads. The general public drove two routes at the Road Research Laboratory test track.

In both cases drivers were shown groups of 2 or 3 digits on a display whose image appeared at the end of a test car's hood. Digits changed in a pseudo random manner. While driving, the participant was asked to call out the numbers shown and "the count." Those responses were recorded on audio tape. At various times the driver was cued to observe the speedometer and then resume counting. The gap in the counting task is a measure of the time to read the speedometer.

Participants drove at 5 speeds (20, 30, 35, 40, 45 mph) as directed by an experimenter. A total of 64 reading times were recorded for each person.

Shown in Table 13 are the results. While the article summarizing the study does not provide the results from significance tests, the error bars shown in the figure in it suggest there were no significant differences between the two numeric displays of the HUD moving scale display and the panel-mounted dial. These results must be viewed with some caution as it appeared that different groups of people saw each display, so that display differences are confounded with differences between test participants.

Other Studies of Speedometers

In addition to the experiments described previously, several others have examined more specific details pertaining to the design of numeric speedometers. Discussion of them and other studies concerning the legibility of numeric displays (Moriyama, Kuroyama, and Shinkai, 1981; Terada, Akeyoshi, and Kadoo, 1982; Yamaguchi, Kishino, and Dorris, 1982) are described in the final legibility report.

Table 13, Mean Reading Times from Armour (1984)

Display Type	Location	Subjects	Reading Time (sec)
moving scale	HUD	police	.85
numeric	HUD	public	.95
numeric (25mm)	panel	------	1.04
numeric (6.4mm)	panel	------	1.10
moving scale	HUD	public	1.56
moving pointer (circular)	panel	------	1.62
moving pointer (linear)	panel	------	2.07

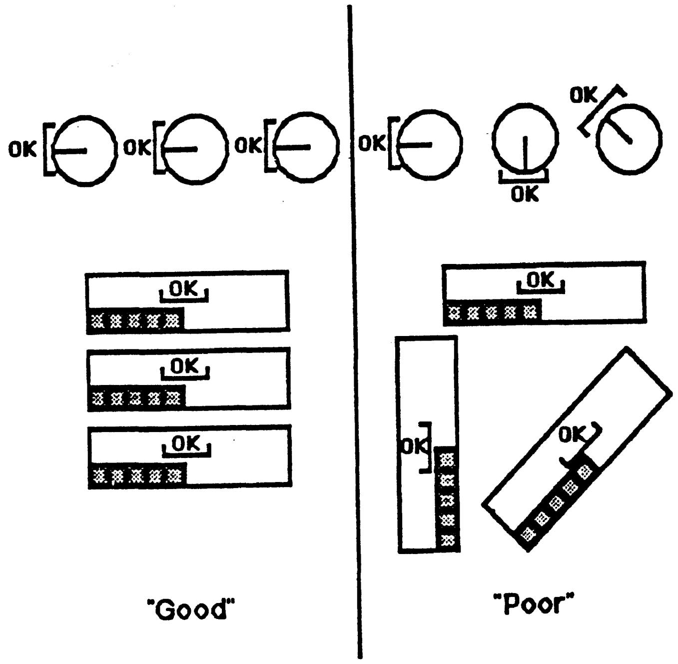
Summary

It seems apparent from the literature that numeric speedometers are easier to read (can be read more rapidly and accurately) than moving pointer designs. The critical evidence is the work of Galer and Simmonds, who examined several designs in a simple lab test, using a simulator, and on the road. The size of the sample (400 drivers) and other factors leave little room to doubt their recommendation.

Surprising in its absence are studies of moving pointer speedometers. While numeric displays are easier to read, they also cost more to make and, as a consequence, mechanical or electro-mechanical moving pointer displays are likely to continue to be fitted in cars in the future. While the general human factors literature offers insight as to how they should be designed, there is still the need for some empiric studies of the readability of common design alternatives.

DOES POINTER ALIGNMENT HELP CHECK READING?

A commonly accepted human factors principle is that a collection of moving pointer displays should be arranged so their pointers are aligned when all displays are showing normal values. This facilitates check reading. Figure 22 (from Green, 1988a) illustrates the idea.



Source: Green (1988b)

Figure 22. Examples of the Pointer Alignment Principle

With regard to that principle, this section addresses the following questions:

- Are there human performance data to support this principle for check reading?
- What normal orientation of the pointer leads to the best check reading performance?
- What else can be done to facilitate check reading?
- Does the principle hold for linear scales as well as for dials?

Pointer alignment was an established principle long before the experimental evidence existed to support it. It was fairly common during WWII for flight engineers to disassemble the engine panel for multi-engine aircraft and re-orient the dials so the pointers were aligned when all displays showed normal values. Often this meant the numbers on the scales were sideways or upside down. Such field modifications are still common today. For example, in race cars, the standard procedure is to mount dials with the pointers normally up, even though the numbers may be inverted. In power plants one often finds similar modifications to align pointers.

Finally, many have argued for pointer alignment based on principles from Gestalt Psychology. In brief, an instrument panel on which the pointers are not aligned when all is normal violates the Principle of Good Form. Simple, symmetrical objects are perceived more readily than those not so configured.

Warrick and Grether (1948)

The classic study of pointer alignment, Warrick and Grether (1948), examined which orientation was best. (See also White, Warrick, and Grether, 1953 for a description of the first experiment.) It was proposed that a 3 or 9 o'clock alignment would be best because the horizontal orientation of the pointers was compatible with reading from left to right.

The same equipment was used for three experiments. A 4x4 array of 1-3/4 inch moving-pointer dials was mounted in a flight simulator (Link trainer). Pointers could be aligned at 3, 9, or 12 o'clock. The instruments were covered by a roller shade that was released when each trial began. Also provided was a compatible 4x4 array of toggle switches the respondent used to identify the dials which were out-of-tolerance, and a single toggle switch to the left of the array used to end a trial.

Each of the 12 men (who participated in all three experiments) saw all of the test displays. The participants had little prior experience with the displays.

In the first experiment, after several practice trials, they searched for the deviated pointer in each array (off by 30 degrees), moved a toggle switch to correct the error, and then called out the nature of the error ("too little" or "too much"). There were 10 test trials per pointer orientation. (In the second and third experiments they also called out "ok" if no pointers deviated from normal.)

In the first experiment, pointers at 9 o'clock were read significantly faster than those at 12 or 3 o'clock (1.96 vs. 2.23 vs. 2.53 s), and more accurately as well (3.3 vs 4.2 vs. 15.8% errors). Limited test statistics were provided.

In the second experiment either 0, 1, or 2 pointers deviated from normal. In addition, all 16 pointers could be aligned at 3, 9 or 12 o'clock, as before, or each row of pointers could be aligned at angles 30 degrees from those positions with every other row having a different alignment.

As shown in Table 14, increasing the number of deviated pointers increased response time. (Error data as a function of the number of deviated pointers were not reported.) Further, there was a huge difference between pointers being aligned at 90 degree increments and those off the axes. (Both response times and errors doubled.) However, the advantage of the 9 over 3 and 12 o'clock positions found in the first experiment did not occur in the second. In experiment 2, the display reading (too high a value) and the associated switch action (move the switch down) were incompatible for the 9 o'clock position, placing that combination at a disadvantage relative to others.

Table 14. Time and Errors for Check Reading from Warrick and Grether (1948)

	-- Mean RT (secs) --				Errors (%)
# Ptrs Out ->	0	1	2	Mean	
** Position **					
(90 degree alignments)					
9 o'clock	1.87	3.10	3.70	3.01	7.5
12 o'clock	1.42	2.69	3.40	2.63	7.5
3 o'clock	1.53	3.36	3.82	3.03	10.4
Mean	1.60	3.05	3.64	2.89	8.5
(Off by 30 degrees)					
9 o'clock	3.22	5.15	6.23	5.06	15.8
12 o'clock	2.96	5.19	6.14	4.96	22.6
3 o'clock	3.27	5.50	6.81	5.42	14.9
Mean	3.15	5.28	6.39	5.14	17.8

The importance of compatibility between the pointer motion and the associated response was specifically examined in the third experiment for pointers aligned at 9, 12, and 3 o'clock. Only 6 people were tested. Participants moved the switch up when the value was too high.

There were no statistically significant differences in either the time (3.84, 3.44, 4.15 s) or errors (23.2, 31.5, 35.8%) between 9, 12, or 3 o'clock, though the trend was for 3 o'clock to be worst. Trends concerning the effects of the number of deviated pointers were similar to experiment 2 (means of 2.26, 3.96, and 4.73 seconds respectively).

Thus, these experiments, involving real displays in a realistic task, show that aligning the pointers for moving pointer dials, when all dials show normal values, significantly reduces check reading time and errors. The particular 90 degree orientation at which that occurs has only a very small effect on performance. The advantages of particular orientations are most likely artifacts of the test procedure (where particular pointer orientations are more compatible with certain switch actions than others).

Senders (1952)

Twenty-four college students viewed arrays of up to 45 oil pressure dials using a tachistoscope. The dials were exposed as long as they pressed a button. Participants then identified the dial with the pointer not aligned at 9 o'clock. Dials were yellow on black, 1-3/4 inches in diameter, and spaced 2 inches apart. Two lighting conditions were examined (1.5 ft-l white, .003 ft-l red).

Shown in Figure 23 is the mean time to check the display as a function of the number of dials. The relationship is clearly linear. For errors the pattern is quite similar with almost no errors being made for 1 dial and 200 for 45.

In the second experiment 8 college students from the first experiment again check read dials, but in this case either dials were aligned at 9 o'clock or a red band (random in location) was provided to indicate normal. On average, the red-banded dials took 18 times longer to read than those for which pointer alignment cues were provided. (See Figure 24.)

In the third experiment 16 college students check read 4 dials by themselves or surrounded by 36 unused dials. Normal was indicated by either the 9 o'clock position, or by a red band. While there again was a difference due to the coding scheme (alignment was better), the mere presence of the unused dials had no effect on performance in this task.

Johnsgard (1953)

To see if patterns of pointers other than those examined by Warrick and Grether (1948) might lead to good check reading performance, Johnsgard (1953) examined those shown in Figure 25. Forty-eight college students, tested in pairs, were shown slides of simulated dial arrays (actually arrays of circles) in which 0 to 6 pointers deviated 15 to 180 degrees from normal. The 72 slides (18 per configuration) were shown for 1/2 second each, a typical fixation duration for a pilot looking at a single display, but not typical for looking at an array of 16 displays. Students checked off the dials with deviating pointers on a response sheet. The viewing distance was 50 inches, slightly greater than the 28 inches commonly used for instrument panel viewing.

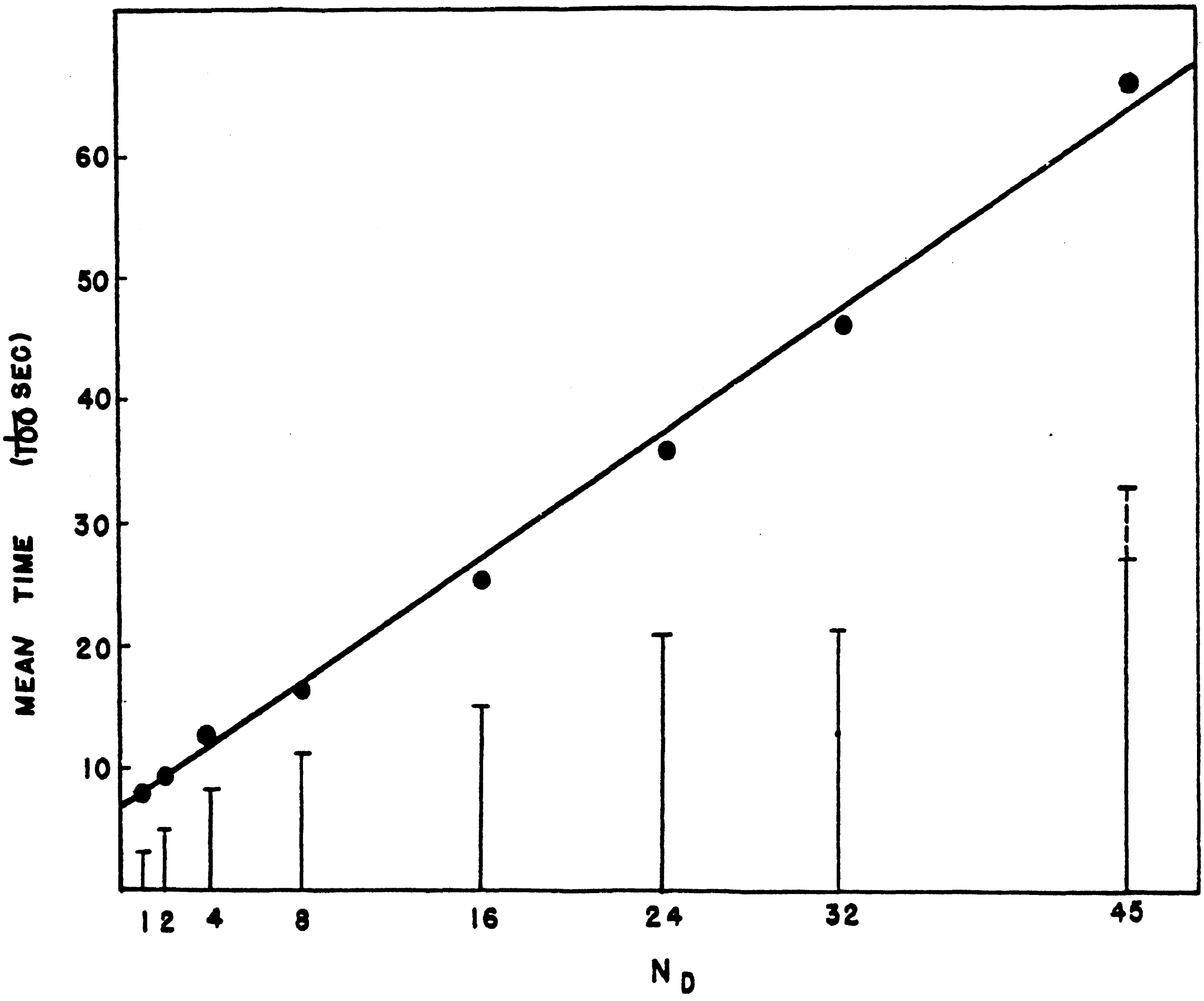
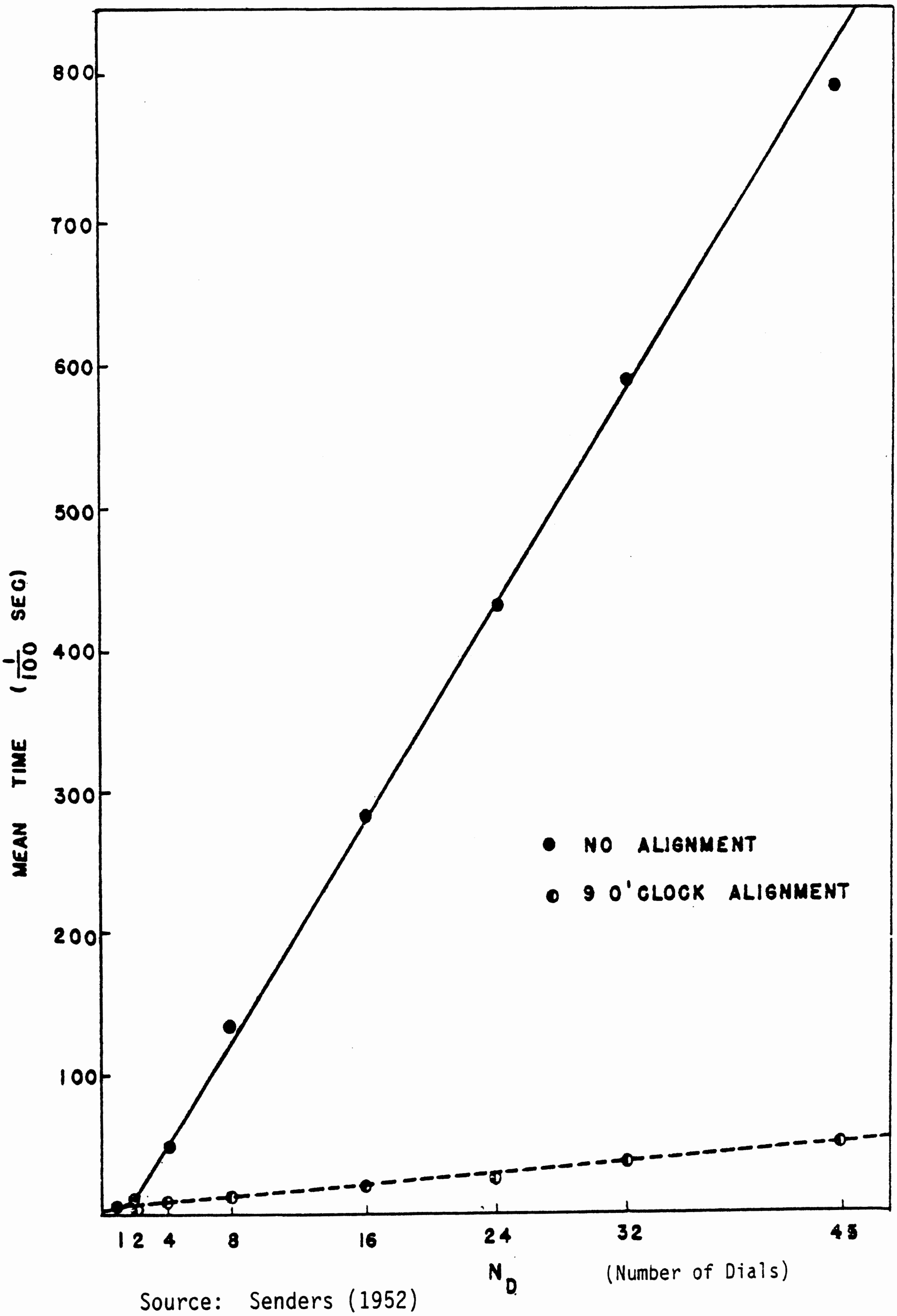


Figure 23, Reading Time As a Function of the Number of Dials
Source: Senders (1952)



Source: Senders (1952)

Figure 24, Pointer Alignment Versus Banding: Reading Time As a Function of the Number of Dials.

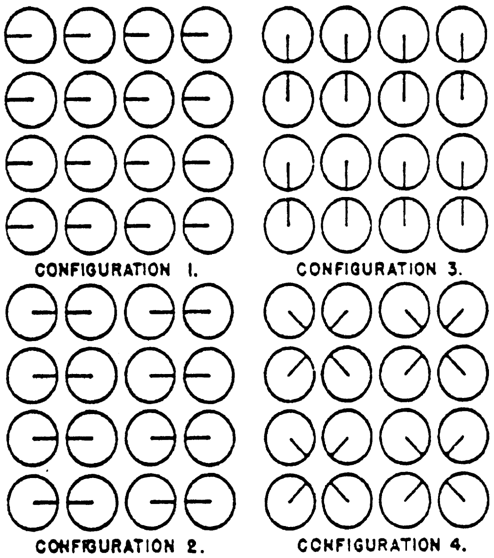


Figure 25. Pointer Configurations Examined by Johnsgard (1953)

* All pointers are shown in their normal or correct position.

Shown in Table 15 are the total number of correct responses to each configuration per respondent. In addition, data on percent correct have been calculated and are included as well. As can been seen in that table, people did not do well in spotting deviated pointers in this experiment. In fact, performance was so poor that questions arise about the applicability of the data to practical problems. In part, the poor performance was due to the relatively brief exposure duration.

Table 15. Correct Detections for Various Configurations Reported by Johnsgard (1953)

Configuration	Mean # Correct	% Correct
1. 9 o'clock	30.2	47.9 *
2. cols 1-2, 3-4 grouped	31.5	49.9
3. rows 1-2, 3-4 grouped	34.5	54.7
4. 4 corner groups	17.5	23.0

* Note: The paper talks about 6 groups of 3 slides with 1-6 deviating pointers per slide. There should therefore be (3x1) + (3x2) ... (3x6) or 61 pointers to detect.

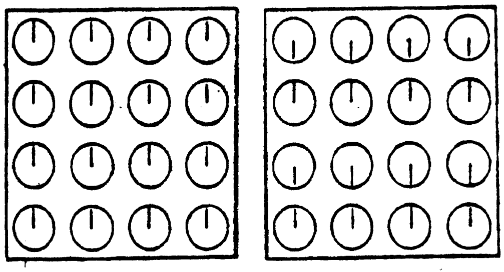
There are also some concerns about the statistical inferences drawn by Johnsgard. He used t-tests to compare condition means when ANOVA and post-hoc multiple comparison tests were more appropriate. (Duncan's test was available at the time (Olson, 1988).) Most likely there are differences between the four-corner configuration and the others, but differences among the others are uncertain.

Ignoring these concerns, the configuration with rows grouped had the fewest errors. While it was not significantly different from grouping rows, Johnsgard reports it was significantly different from the standard 9 o'clock configuration. People did not do well with the four-corner group arrays.

Most conspicuous are the factors that Johnsgard manipulated but did not analyze--the effect of the size of the angle of the deviating pointer and the number of deviating pointers on percent correct.

Ross, Katchmar, and Bell (1955)

Ross, Katchmar, and Bell (1955) describe an experiment similar to that of Johnsgard. Twenty-four college students were shown 69 (5 practice, 64 test) arrays of 4x4 dials with moving pointers. Pointers were either aligned at 12 o'clock, 6 o'clock, or in row pairs (Johnsgard's configuration 2). (See Figure 26.) Projected images of the dials, shown for .2 seconds each, were viewed from 8 feet away (though the visual angle was typical for dials viewed at more typical instrument panel distances). In each array, one dial had a deviated pointer (by 90 degrees). Students marked the out-of-tolerance dial and its direction on a response sheet.



The configurations tested in the experiment. The arrangement on the left is C1 (Uniform Alignment at the 12 o'clock position). The arrangement on the right is C2 (Pointer Symmetry), where Row 1 and Row 3 are set at 6 o'clock and Rows 2 and 4 at 12 o'clock. C3 (not shown) is the same as C1 except that the pointers are set at 6 o'clock. The configurations do not show the single deviant pointer, which varied from trial to trial.

Figure 26. Configurations Tested by Ross et al. (1955)

In the first experiment, the first two configurations were examined in an order counterbalanced across subjects. There were no overall statistically significant differences, though there were slightly fewer positional (10.0 vs. 11.0) and directional (6.3 vs. 7.3) errors per person for the 12 o'clock configuration.

In experiment 2, configurations 2 and 3 (pointer symmetry and 6 o'clock) were examined by the same people. There were two test sessions instead of the one in the previous experiment. The average number of positional errors was significantly less for configuration 3 than 2 (5.6 vs. 7.1), but the difference in directional errors was not significant (3.7 vs. 4.3).

Thus, Ross et al. show that pointer symmetry provides performance gains over just uniform alignment. They argue that they were more likely to find differences than other researchers because the shorter exposure duration they chose (making the task more difficult) was more likely to identify significant differences.

Dashevsky (1964a)

A number of studies have examined how to enhance the benefits of pointer alignment cues. In Dashevsky's first experiment, carried out to develop the test method, 25 people participated. They sat in a darkened room 8-10 feet from a projection screen and viewed it either head on or peripherally, 90 degrees to the left or right. Arrays of circles (4x4) simulating dials were shown. Deviated pointers (maximum of one per array) were shown on 1/2 of the trials. When presented, they deviated by 90 or 270 degrees from normal. Normal was either 9 or 12 o'clock. Participants marked on a sheet if the pointers were aligned.

The effect of pointer position (9 vs 12 o'clock) was not significant though there were fewer errors for the 12 o'clock position (e.g., for the head on condition, 24.5 vs 21.5 errors).

In the second experiment six designs were examined (Figure 27). It was thought that extending the pointer would make it easier to detect alignment differences. Of the 20 dial arrays shown for each configuration, 15 had one deviated pointer (90, 180, 270 degrees). Arrays were shown for .5 seconds each.

Shown in Table 16 are the total errors for each design. There were major differences between configurations. The critical result was that extending pointers reduced reading errors by 85%, a statistically significant result. It is difficult to say why real display systems have not been designed with this feature. As a note of caution, the angular deviations examined in this study were quite large. It would be useful to know what the performance advantages were of extended pointers when the unacceptable deviations from normal were smaller and the pointers showing normal varied slightly.

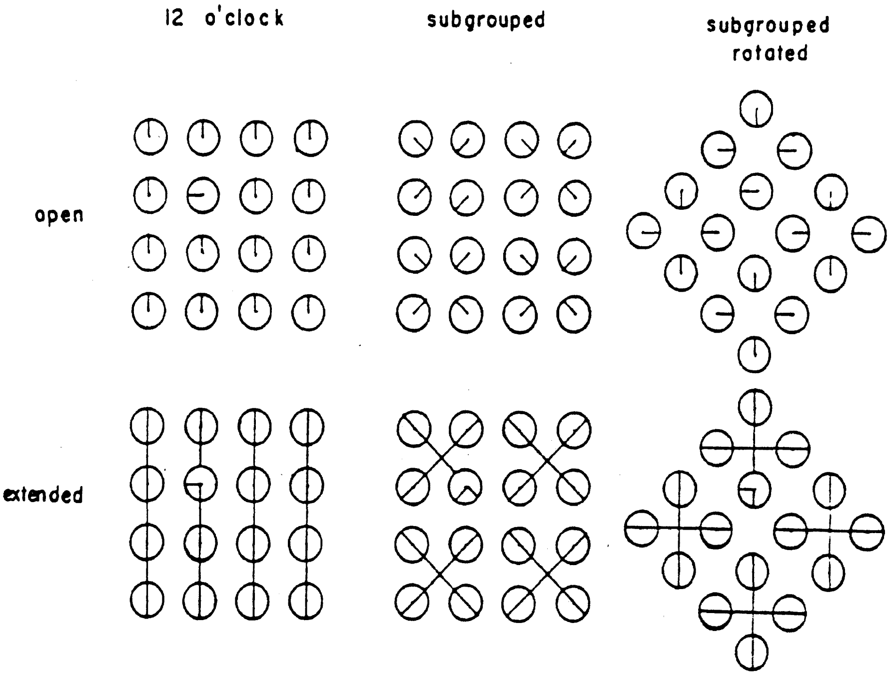


Figure 27. Designs Examined by Dashevsky (1964a)

Table 16. Total Errors from Dashevsky (1964)

Pointer	Configuration 12 o'clock	 Subgrouped	 Subgrouped & Rotated
Open	53	193	201
Extended	8	15	41

Dashevsky (1964b) extended these results to other tasks and related display formats. Those formats tested are shown in Figure 28. The test procedure was similar to that used previously. People were shown 20 test slides of 4x4 arrays of simulated meters exposed for 1/2 second. No more than one pointer per array deviated from normal. Subjects marked a form to indicate which pointer it was.

Forty people were tested. It appears that each person saw only one of the four test designs.

In general, people did significantly better with the qualitative version than quantitative, and better with the semicircular than the circular meters. (See Table 17.) One might have expected the quantitative display to be easier to check read. It may be that the size of the arc on the quantitative display made detecting the pointer difficult and redesigning it could enhance performance.

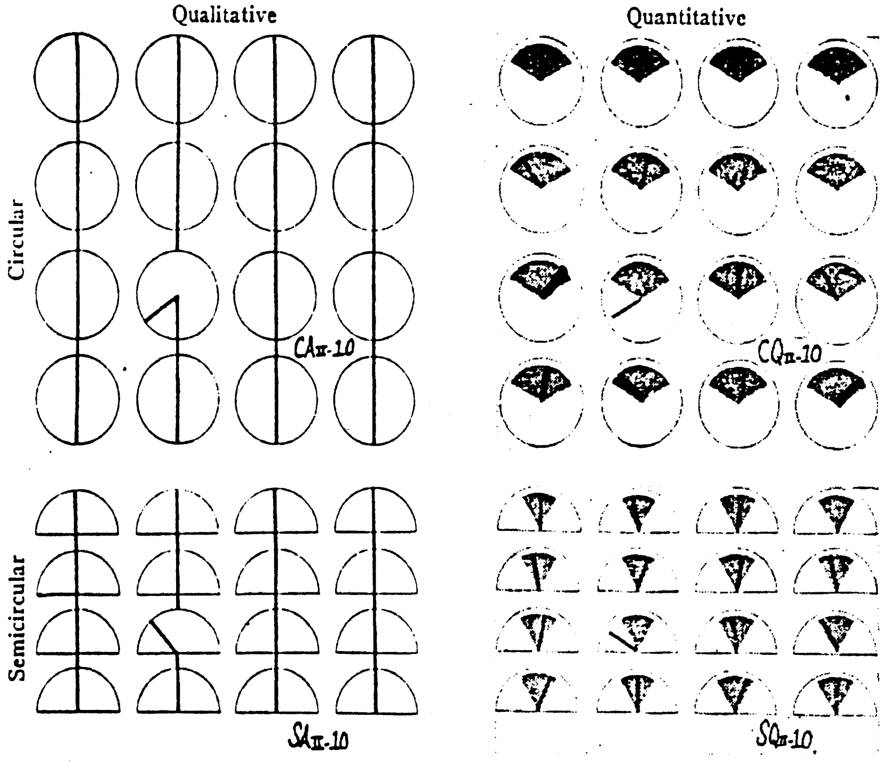


Figure 28. Displays Tested by Dashevsky (1964b)

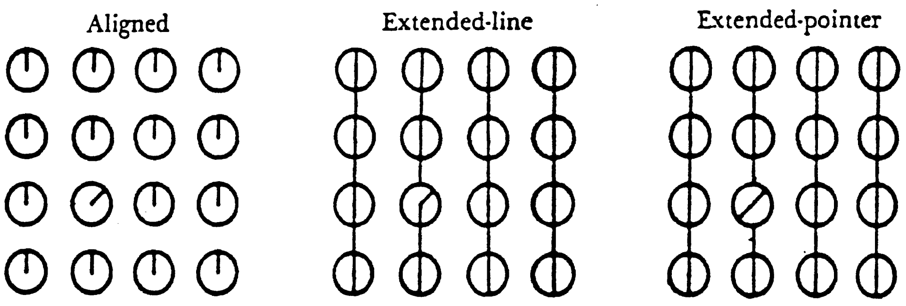
Table 17. Error Percentages and Number of Errors in Dashevsky (1964b).

Display Shape	----- Intended Task ------- Qualitative	Quantitative
Circular	2.75% (4.5)	5.75% (11.5)
Semicircular	0.25% (0.5)	2.25% (4.5)

Readers should view these numbers with some caution. The differences between displays were confounded with differences between groups of test subjects. Further, the error rates for all configurations were very low and differences between them represent only a few errors. (There were 200 responses per configuration.) Those who look carefully at the numbers will find that the number of errors are not integers. The scoring procedure used allowed for "1/2 errors" and there are concerns that the procedure was arbitrary.

Oatman (1964b)

Oatman (1964b) (see also Oatman, 1964a) describes a continuation of the extended-pointer display study. (See Figure 29.) Thirty-four enlisted Army men were shown 3 sets of 17 black and white slides exposed for .04 seconds each. Each set of 17 had 16 slides in which a different single pointer deviated from normal (by 35 degrees) and one where all were normal. The 4x4 arrays of simulated 3-inch dials were viewed from a distance of 50 inches. Each set was seen by every participant four times.



Source: Oatman (1964b)

Figure 29. Displays Examined in First Oatman Experiment

The task was the same as that used by Dashevsky. Soldiers indicated if the slide had a dial with a deviated pointer (detection task) and identified the dial (location task).

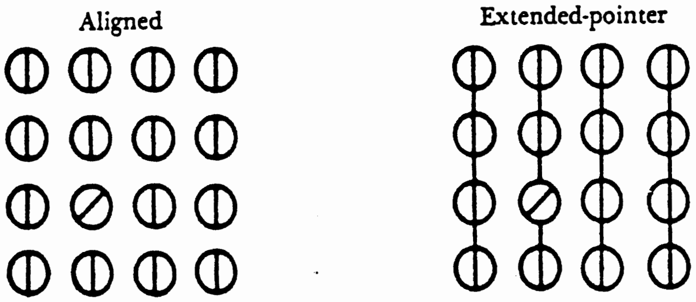
Shown in Table 18 are the percent correct data derived from Oatman's data. Soldiers did significantly better with the extended pointer display. There were no differences between the other two types of displays. A detailed analysis of the location of the errors showed that the largest number occurred in the lower right corner, a finding in agreement with previous studies.

Table 18. Mean Percent Correct Detection and Location Scores Derived from Oatman (1964b)

Display	Detection	Location	Mean
Aligned	93.9	82.2	88.1
Extended-line	93.4	84.5	89.0
Extended-pointer	98.9	90.2	94.6

Oatman (1965a, b)

A second experiment (see also Oatman, 1964b) was conducted to determine if the initial extended pointer design did well because of the pointer alone or because of the pointer the extension line. Figure 30 shows the displays examined.

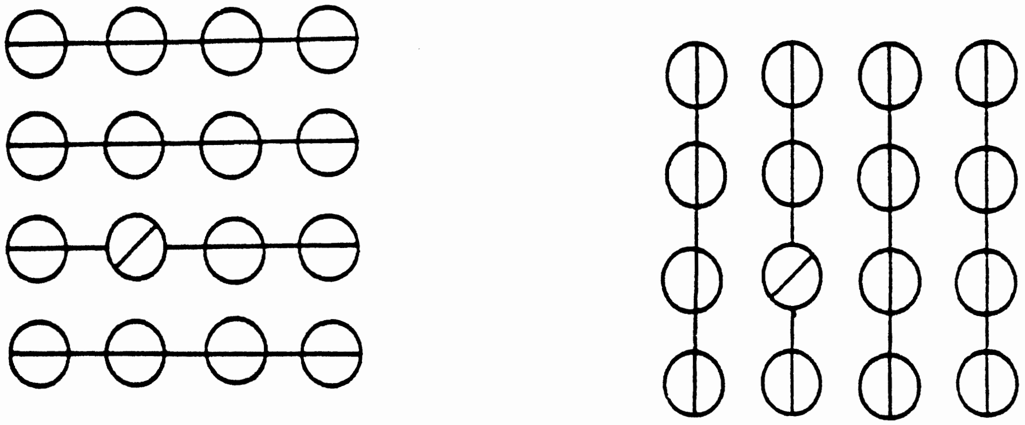


Source: Oatman (1965a)

Figure 30. Displays Examined in Second Oatman Experiment

Thirty enlisted Army men served in an experiment similar to that just described. No significant differences in detection errors (Aligned=4.53, Extended-pointer=4.58) or location errors (3.93 vs 3.99) were found. Hence, the key to improving the check reading of pointer displays is extending the length of the pointer. Adding extension lines between displays has no effect on performance.

As a final check of the effectiveness of extended-pointer displays, 9 and 12 o'clock pairs similar to those in Figure 31 were examined. The detection/location tasks described previously were carried out by 30 Army men. No significant differences in detection scores (12 o'clock=4.62, 9 o'clock=4.70) or location scores (3.96 vs. 3.86) were found. This agrees with previous research (e.g., Warrick and Grether, 1948) showing that the position around which pointers are aligned has no effect on performance.



9-O'clock Extended-Pointer Display

12-O'clock Extended-Pointer Display

Figure 31. Extended-pointer Dials Examined by Oatman (1965a)

Bauer, Cassatt, Corona, and Warhurst (1966)

The benefits of pointer alignment have also been examined for linear displays. In their first experiment, Bauer, Cassatt, Corona, and Warhurst (1966) showed 36 pilots aircraft instrument panels similar to that in Figure 32. There were three basic arrangements--vertical scales, horizontal scales, and mixed scales (as shown in Figure 33) with 1/3 of the participants seeing each of the basic arrangements. Each group saw two versions of each configuration, uniform (shown in Figure 33) and non-uniform (Figure 32). Uniform displays are apparently those for which the digit font, digit placement, tick mark size and intervals, and other characteristics are consistent across displays. According to these figures, there were 16 linear scale displays on each panel. It is not known how many pointers were deviated per condition or whether pointers were aligned when normal (though it is presumed they were). This makes interpretation of the results quite difficult.

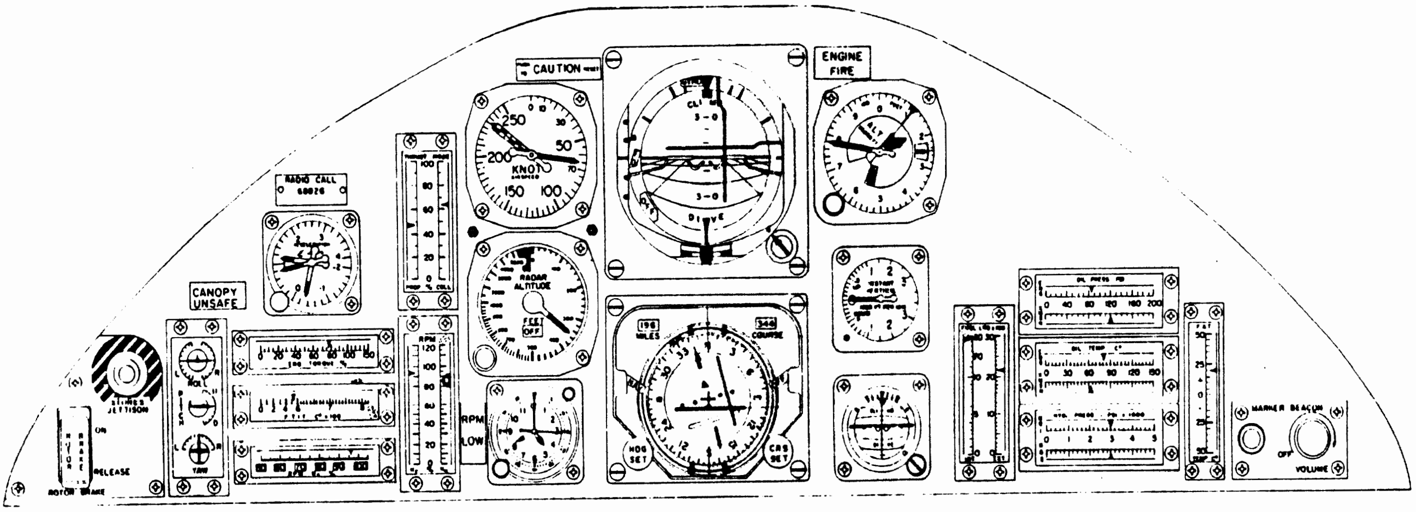
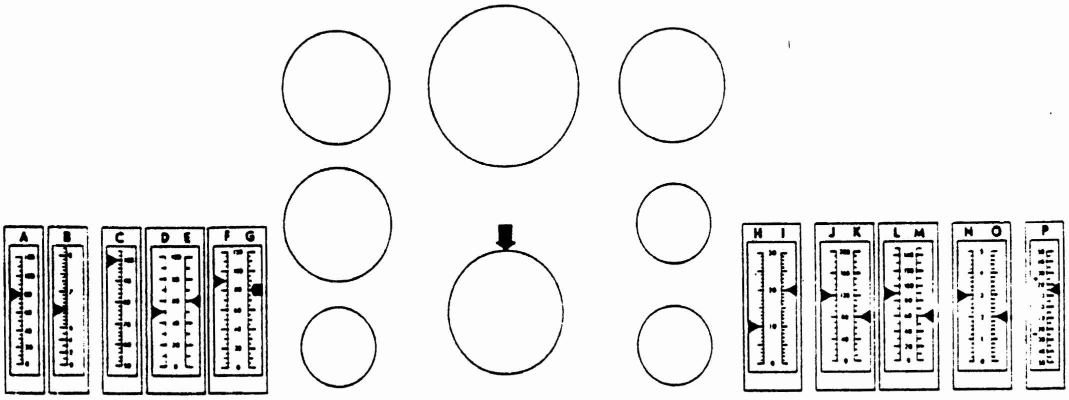
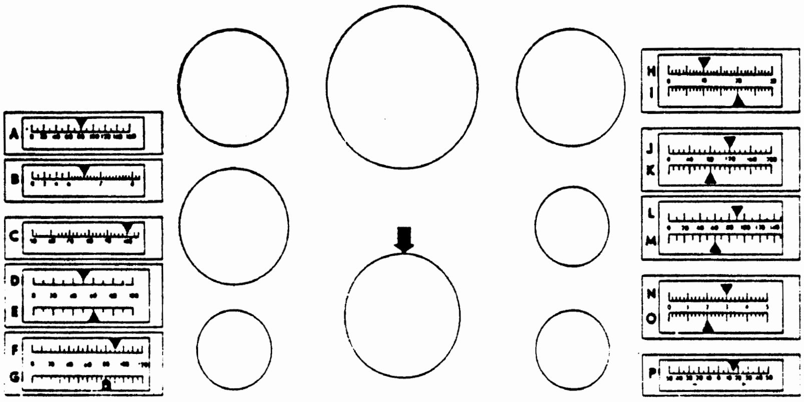


Figure 32. Typical Instrument Panel Examined by Bauer et al. (1966)

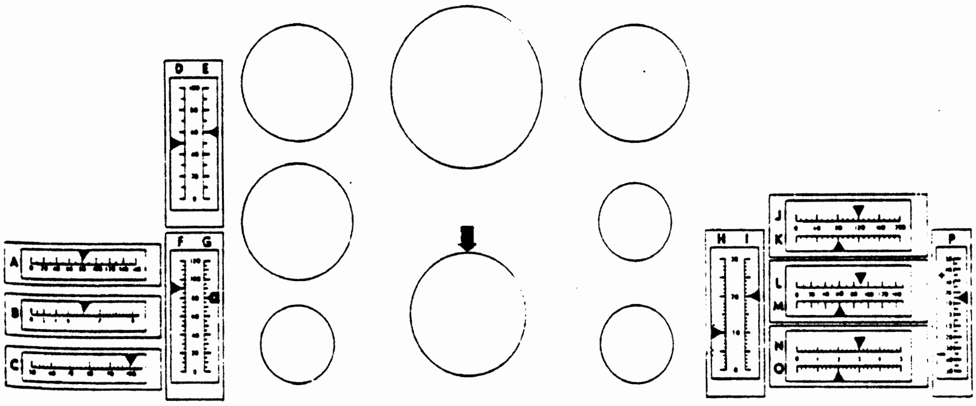
After being shown a demonstration slide, each person was shown a standard slide (depicting the normal values) and then four test slides. Participants controlled the exposure duration and tried to keep it as brief as possible. For each slide the participant read the compass heading (the central display), identified by letter those linear displays out-of-tolerance, and read out (presumably aloud) the values displayed for them. Thus, as with many of the previous studies, there were two classes of responses, those associated with detecting errors, and those associated with identifying the magnitude or direction of the error.



Vertical dial layout; uniform scales.



Horizontal dial layout; uniform scales.



Mixed dial layout; uniform scales.

Figure 33. Scale Designs Examined by Bauer et al. (1966)

A major advantage of the Bauer et al. work over previous studies is that it begins to structure the check reading of gauges as a signal detection problem. The signal-detection perspective is a framework for describing how people and machines classify information. For those unfamiliar with this perspective, Table 19 provides an overview. There are four stimulus-response combinations, two of which represent errors (misses, false alarms) and two which represent positive outcomes (hits, correct rejections). In most studies of gauge reading, only misses are reported. In practical situations, both kinds of errors are important. The advantage of this perspective is that it separates strategies of participants and their biases, from genuine display design differences. For example, one way for a person to never to miss a deviated pointer would be to say that every pointer was deviated, whether or not it actually was. The cost of this strategy is that the false alarm rate would be very high. If a person changes their bias in an experiment, and only one type of error is recorded, the outcome can be misleading. It is usually assumed that biases are fixed.

Table 19. Categories in a Signal Detection Analysis

		Response (What does the person say?) Yes, it is	 No, it isn't
Stimulus	Yes	Hit	Miss
(Is it really deviated?)	No	False Alarm	Correct Rejection

Table 20 shows the detection and other data provided by Bauer et al. Utilizing uniform scales significantly reduced the number of reading errors. Surprisingly, there were also differences due to configuration, with the mixed configuration having significantly more hits, fewer false alarms (higher detection rate) and the shortest detection time. (See Table 21.) This disagrees with what one would expect from the literature on the pointer alignment principle, and Bauer et al. are unsure why it occurred. Readers should bear in mind each display type was seen by a different group of pilots and no effort was made to match groups.

It is unfortunate that Bauer et al. did not take their analysis one step farther and carry out a complete signal detection analysis. Since they don't identify the number of times pointers deviated for each arrangement or the total number of trials, the number of misses and correct rejections cannot be computed. Therefore, one also cannot compute d', the measure of detection performance.

Table 20. Detection and Reading Errors Reported by Bauer et al. (1966)

Scale	Scale Markings: Uniform Hits	FA	Errors*	Non-Uniform Hits	FA	Errors
Vertical	87	30	3	85	26	14
Horizontal	82	38	2	98	19	11
Mixed	110	19	4	100	10	17
Total	279	87	9	283	55	42

*Note: In the Bauer et al. paper the terms "Detections, True-Positive," Detections, False-Positive," and "Readout Errors" are used instead of Hits, FA (False Alarms), and Errors. "Readout Errors" most likely refers to the correct detection deviated pointers (a hit), but then subsequently misreading the value displayed.

Table 21. Mean Time/Detection (seconds) Reported by Bauer et al. (1968)

Scale Configuration	Scale Markings: Uniform	Non-Uniform
Vertical	13.6	15.0
Horizontal	15.1	13.8
Mixed	9.4	11.7

The task in the second experiment was the same as in the first. In that experiment, 16 non-pilots (technical and professional laboratory personnel) were shown two sets of four slides showing uniform vertical scales only, either 3/8 or 5/8 inches apart. (That is assumed to be edge-to-edge separation.) There were no significant differences between the two separations (3 readout errors, 13.4 seconds/detection for the narrow spacing; 9 and 12.3 for the wide). Performance in this experiment was similar to that of the first, suggesting that non-pilots could be used to predict the performance of pilots.

Mital and Ramanan (1985)

The most recent study concerning the merits of 9 and 12 o'clock positions is Mital and Ramanan (1985). (See also Mital and Ramanan, 1986). Fifty college students were shown slides of 4x4 arrays of simulated dials similar to those in Figure 34. Both black on white and white on black displays were examined. Slides were shown for either .25, .5, or .75 seconds. The percentage of deviated dials varied from 1 to 3%. The viewing distance was 1.27 m (50 inches) and the dial diameter 7.62 cm (3 inches).

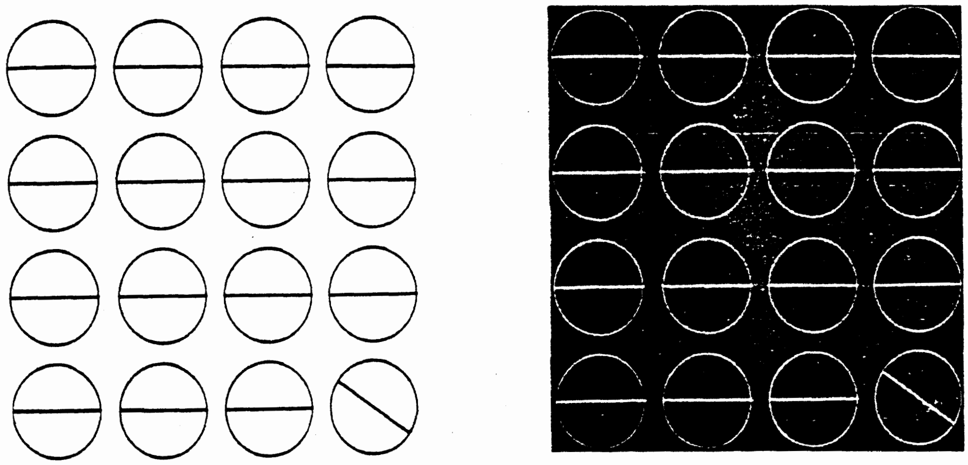


Figure 34. Displays Examined by Mital and Ramanan (1985)

Figure 35 shows the results. In general, people did considerably better at longer exposure durations. They did better with white on a black background for the longest exposure duration. The reverse was true for shorter durations. However, the percentage of deviated dials did not have a significant effect on the error percentages (1%=1.36, 2%=1.42, 3%=1.45). Most importantly, there were no significant differences due to the normal pointer position (9 o'clock=1.36, 12 o'clock=1.47). This may be due to floor effects of the test conditions chosen. (The task is easy enough that people make almost zero errors, performance one cannot improve upon.) Mital and Ramanan note that people preferred the 12 o'clock position and therefore they recommend it, but they do not provide any supporting statistics.

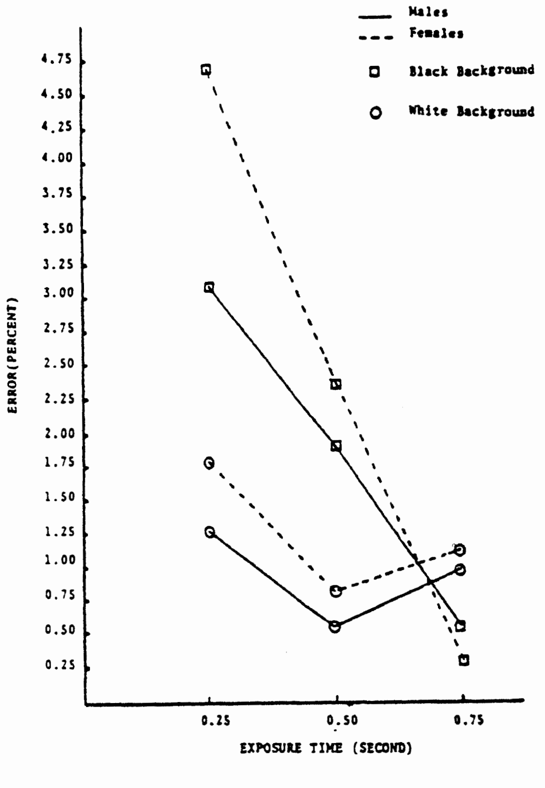


Figure 35. Results from Mital and Ramanan (1985)

Poynter and Czarnomski (1988)

In a well controlled experiment, Poynter and Czarnomski had 19 people look at 216 arrays of items presented for 150 ms each on a video display. They examined 6 display formats (whether a line, number, gap in a circle, or some combination indicated a display was out of tolerance). They also varied the number of displayed items (4, 9, or 16) and the number of targets (1, 2, or 3 deviant display elements). See Figure 36. The participant's task was to identify "off-normal" elements using a mouse.

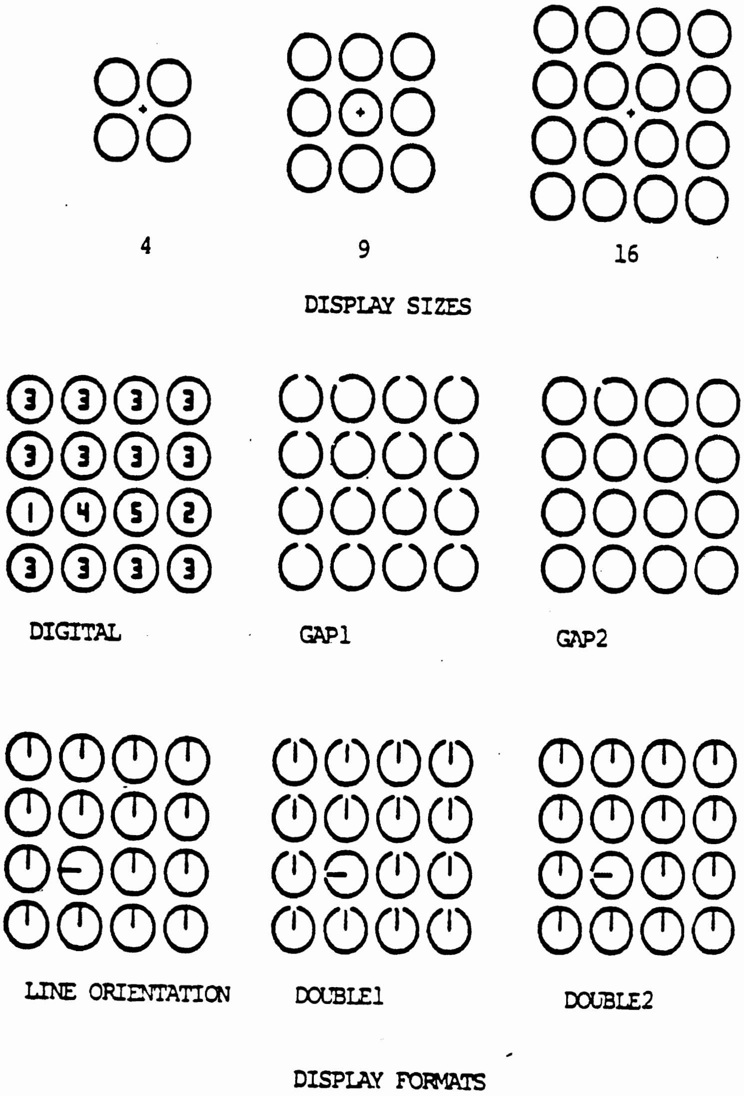


Figure 36, Displays Examined by Poynter and Czarnomski (1988)

Shown in Figure 37 are the results. Performance measures were how likely they were to detect that something was off-normal, to identify ("localize") the item, and to report its value. Error rates tended to be lower for the pointer displays. Those rates went up linearly with the number of display elements for the numeric displays but were relatively unaffected for the pointer displays. In both cases increasing the number of targets led to linear increases in error rates. With regard to the various pointer designs, it appears that the simple pointer display was best followed by the broken circle display (gap2) followed by the rotated circle display (gap1). The combined displays (double1 and double2) were better (led to fewer errors) than the simple pointer display, but the differences were small.

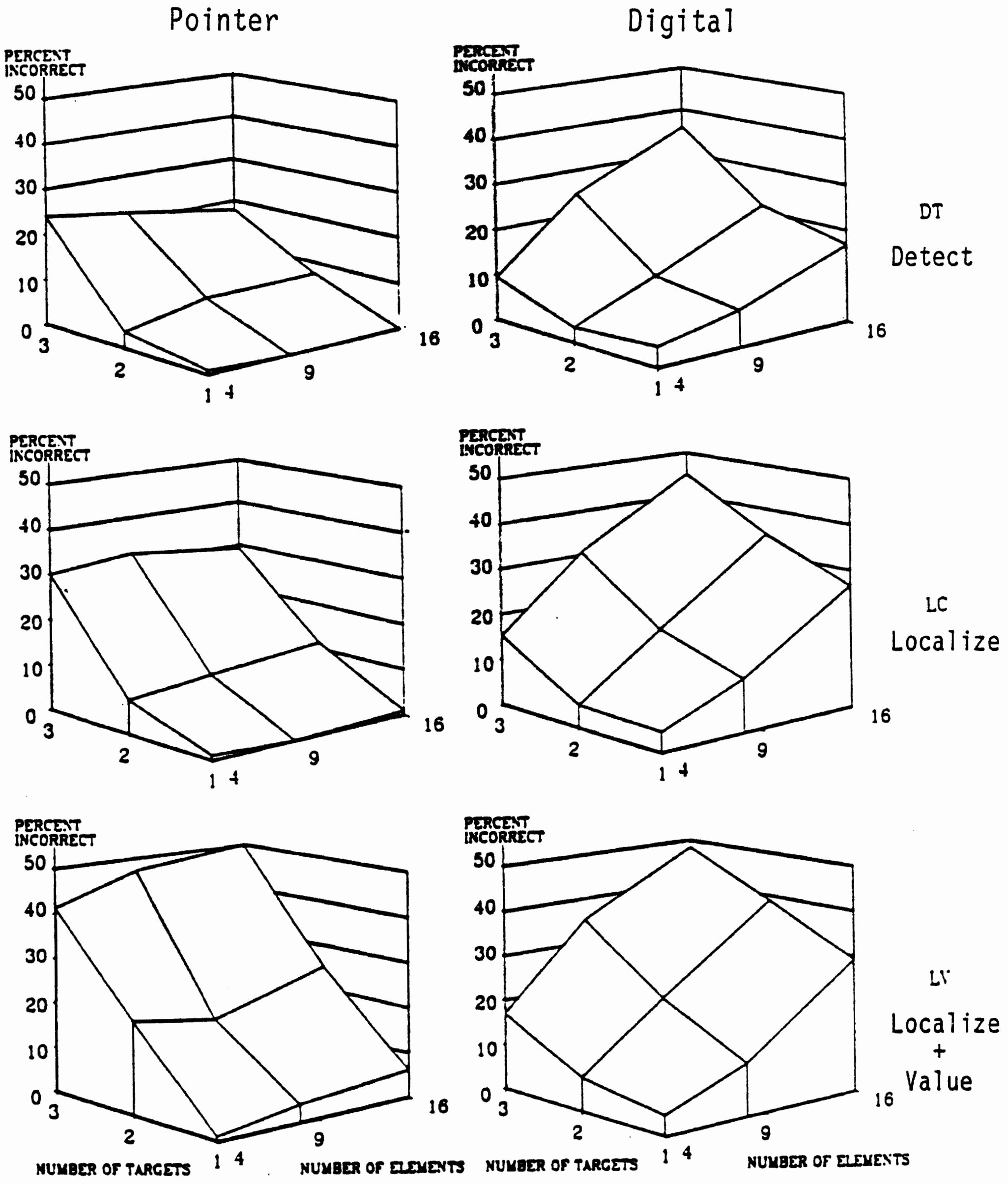


Figure 37, Error Data from Poynter and Czarnomski (1988)

Summary

1. Pointer alignment cues aid the check reading of displays. When compared with reasonably good but not optimized display groups, the reading times and errors may be cut in half.

2. There is no single best position for pointers to be aligned to facilitate check reading, but clearly the 4 cardinal clock positions (3, 6, 9, and 12) are superior to alternatives. Virtually every study covered in this section touches upon the orientation issue.

3. Extending pointers so they cover the entire dial face facilitates check reading.

4. Providing extension lines for pointers to emphasize alignment has not been shown to improve performance in well practiced tasks. However, the author believes that such lines could help drivers realize the pointer alignment principle has been incorporated in an instrument panel. This has not been examined experimentally.

5. While it is believed that the pointer alignment principle extends to linear scales as well as dials, formal experimental evidence doesn't exist. The one experiment that examined this issue did not support the principle. The study, however, contains sufficient flaws that it should be discounted.

How applicable are these recommendations to automobile instrument panel design? While it is true there are no four engine cars and few cars with 16 dials to show engine performance, these principles apply, even if there are only two gauges. Many cars have three (engine temperature, oil pressure, and some measure of the electrical system). As cars increase in complexity, there is reason to believe the number of displays will increase.

There are, however, several issues that should be addressed for automotive applications. First, it would be useful to have data showing the benefits of pointer alignment for linear displays. Clean experimental evidence doesn't exist, though there is good reason to accept the principle. Further, many automotive designers tend to ignore the literature because "it has to do only with airplanes." Should such work be conducted, signal detection analysis methods should be used to examine the error data.

Second, it would also be useful to know what the performance benefits are when the pointer alignment for normal is imperfect. This is often the case in real systems. (It would also be useful to know how detection performance varies with the extent of the deviation. Most experiments have examined only large values (e.g., 90 degrees).

These recommendations given in this section are well supported by the research literature and easy to apply. Furthermore, the recommendations have been around for some time and appear in most common human factors design standards. In spite of this, the number of automobile instrument panels that satisfy even these basic requirements are few.

HOW SHOULD SCALES BE MARKED?

In the late 40's and most of the 50's considerable research was conducted concerning how scales should be marked. Very little work was done afterwards because most of the basic questions were answered. Those questions include:

What is the relative importance of the scale marks?

How far apart should the marked intervals be?

In what increments should scales be marked?

How large should scale marks be?

Are there instances where nonstandard marking schemes (staircase, log, etc.) can enhance readability?

Several terms are used throughout this section to refer to scale elements. The scale range is obviously the difference between the smallest and largest value that can be shown. The size of the marked interval refers to points on the scale that have numbers on the scale next to them. So for 0 to 10 scale on which only "0" and "10" appeared, the marked interval would be 10. The size of the called interval refers to how accurately a scale is read. If a 0 to 10 scale is read to the nearest unit, then the size of the called interval is 1.

Vernon (1946)

This difficult to obtain and poorly written report describes a series of early studies of dial reading. In the first experiment, 24 RAF pilots were shown 4 sets of 20 scales. In each set, half were circular and half were horizontal. Sets varied in terms of what the pointer looked like and how far it was from the scale graduations (Set A), the integers used to label intervals (1's, 2's, 4's, etc., Set B), the decimals used to label intervals (.1's, .2's, etc., Set C), and the number of marks between labeled scale divisions (2, 4, etc., Set D). Details concerning the number of each variation included in each set are not provided.

Each scale was exposed for 2 seconds in a tachistoscope. The pilot read the scale as accurately as possible.

There were no differences between straight and circular scales except when the pointer was more than .8 inches from the scale graduation and readings without decimal points were made. Gaps of 0.5 inches or less are recommended by Vernon. Apparently for short pointers, alignment is more difficult for circular scales. (Error data as a function of the gap are not provided). This outcome may also be due to inconsistencies between the test conditions (differences in the minor scale marks). When the optimal dimensions are chosen for the pointer, there appear to be no differences between scale types.

Shown in Table 22 are percentage of scales misread as a function of the scale interval markings. People do best with scales numbered using factors of 10, but have considerable difficulty if scales are labeled with values less than 1, even if they are even multiples of 10.

Table 22, Errors (%) Versus Scale Numbering Intervals

Scale	Interval			
	100, 10, 1, .1, .01, .001	20, 2, .2 .02, .002	250, 25, .125, .025	50, 5, .5, .05, .005
Straight	15.6	32.3	80.2	56.3
Circular	11.7	80.2	43.1	19.8

Source: Vernon (1946)

Finally, there is some mention of the number of marks between labeled intervals, but the conclusions are unclear.

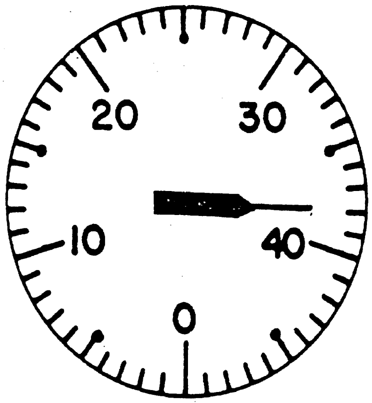
In the second experiment 9 people read sets of 11 dials in rapid succession. Each set was comprised of 3 banks. In one bank all dials were labelled in the same way. In the others, the consistency of the labelling schemes varied in a disorderly manner. Each set was read several times over a two-day period. The exact number of readings depended on the sample of participants.

While the report does show the total times to read each set of 11 four times, reading times for the individual banks are not provided. Vernon makes the point that it was easier to read the dials when successive dials used the same numbering scheme, but no supporting data are provided.

The third experiment also examined the role of the consistency of a set on reading performance. The dials used were similar to the previous experiment. Eighteen people repeatedly read the values shown on 9 dials. In general, the time to read the set when the scales were labeled the same was almost half the time required when they were mixed. Thus, in several experiments, this study shows that the consistent marking of a series of scales, and labelling individual scales with numbers that are greater than one and even multiples of 10, can enhance performance. However, the description of how these conclusions were formed could have been much better.

Christensen (1948)

A common mistake that people make when reading dials is a reversal error. An example appears in Figure 38.

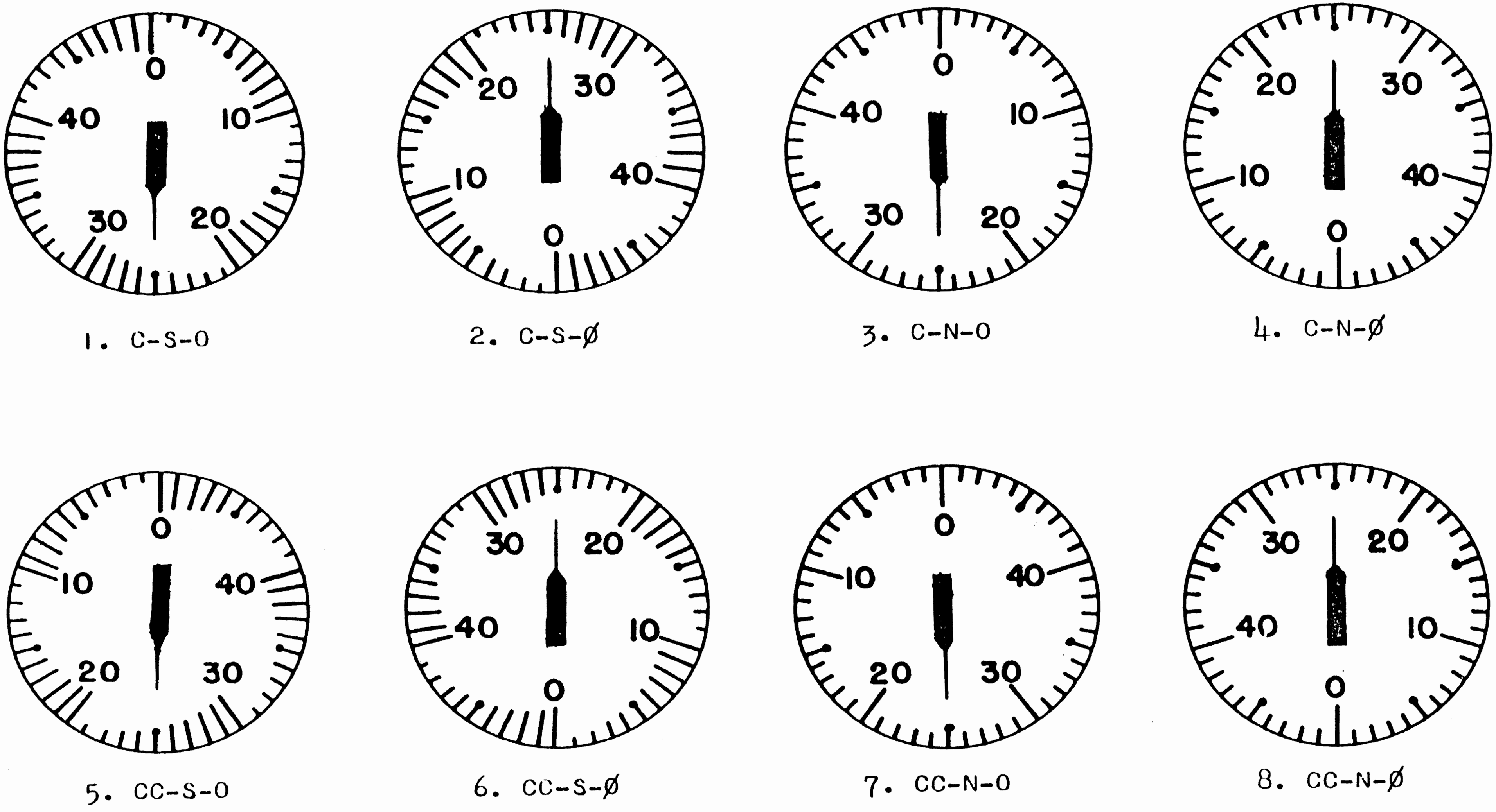


"38" misread as "42" (Source: Christensen, 1948)

Figure 38, An Example of a Reversal Error

It was thought that varying the size of the minor marks in a staircase fashion might make such errors less common. Christensen showed 20 pilots and 33 college students cards with eight dials on them. (See Figure 39 for a sample.) Their task was to write down the values shown as rapidly as possible and to record the time required to complete each card. Dials varied in terms of where zero was located, whether they were numbered in a clockwise or counterclockwise manner, and the minor scale mark design (normal or staircase).

The overall error rate was about 10%. About 1/2 of the errors were reversal errors, which the staircase design was intended to reduce. Reversal error rates were 5.06 and 3.70% for college students and pilots respectively for the staircase design, and 6.52 and 4.25% for the normal design. The difference was not statistically significant. For this and other reasons, a change in design practice was not recommended by Christensen. The numbering direction did have a major effect on performance. There were 2-1/2 times more reversal errors for counterclockwise dials. On the other hand, the location of the zero point had no effect.



1. C-S-O

2. C-S-0

3. C-N-O

4. C-N-0

5. CC-S-O

6. CC-S-0

7. CC-N-O

8. CC-N-0

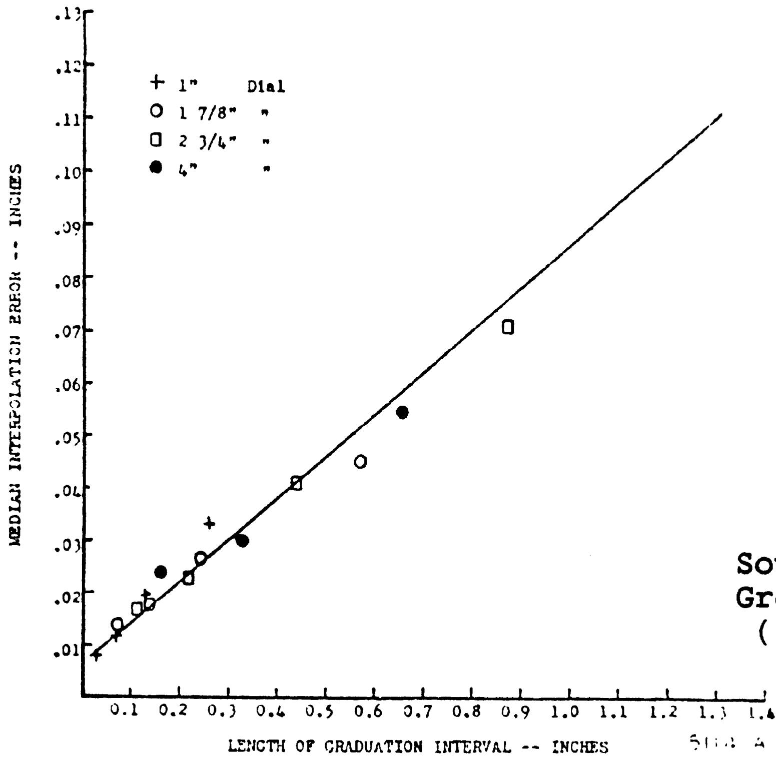
Figure 39, Dials Examined by Christensen (1948)

Grether and Williams (1949)

In this study (see also Grether, 1947) subjects were shown single simulated instrument dials either 1, 1-7/8, 2-3/4, or 4 inches in diameter. The two intermediate sizes were used in aircraft at the time. Graduation marks were spaced every 5, 10, 20, or 40 degrees. Dials were viewed under simulated daylight (45 ft-c) or nighttime conditions at a distance of 30 inches.

Eighty male college students with 20/20 vision or better participated. They read the dial as well as they could, interpolating to the nearest tenth of a mark. Each responded 80 times seeing only 1/4 of the set of dials and under either daylight or nighttime conditions. On each trial the experimenter set the pointer to a random position, opened the shutter, and closed it when the participant started to read the dial.

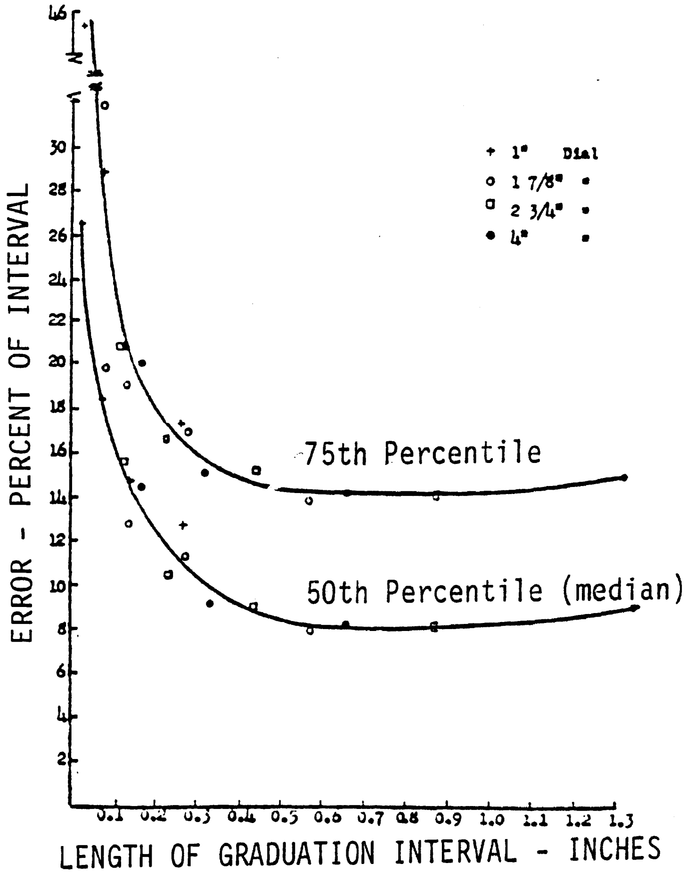
In general, differences in performance between the daytime and nighttime conditions were small. Shown in Figure 40 is the median interpolation error as a function of the length of the graduation interval (the length of the curved baseline around the dial). Notice that the size of the errors are proportional of the visual angle with respect to the observer.



Source:
Grether and Williams
(1949)

Figure 40, Absolute Error of Interpolation As a Function of Graduation Interval Length

Figure 41 shows the relative error as a function of graduation interval length. This is a typical Weber function (plot of intensity (I) versus threshold ratio (delta I/I). Notice that the relative accuracy of interpolation is nearly constant for intervals above 1/2 inch. These data have been used to argue for having graduation marks about 1/2 inch apart where interpolation is required (for gauges viewed at 30 inches).



Source:
Grether and Williams
(1949)

Figure 41, Relative Error of Interpolation As a Function of Graduation Interval Length

There were no consistent relationships between the dial diameter, the graduation interval, and the time to read dials, though readings for the nighttime condition did take slightly longer. The lack of relationships is most likely due to the crude method used to measure response time.

Kappauf and Smith (1950)

Twenty high school students, all with at least 20/20 vision were shown cards on which 12 dials appeared. They were timed as they read the center 10 dials "to the nearest unit." Fifty or 60 readings were made for each dial. A total of 34,400 data points were collected with each subject participating in six one-hour sessions. This impressive sample size is about an order of magnitude larger than is found in a typical experiment.

Dials were white on black. The character luminance was 3 ft-l. Fifteen designs were examined. (See Figure 42.) Variations included the number of minor markings (every unit, every two units, five units, or ten units), numbered increments (10's, 20's, 40's and 100's), and the scale range. Two sizes were examined, 1.4 and 2.8 inch diameter.

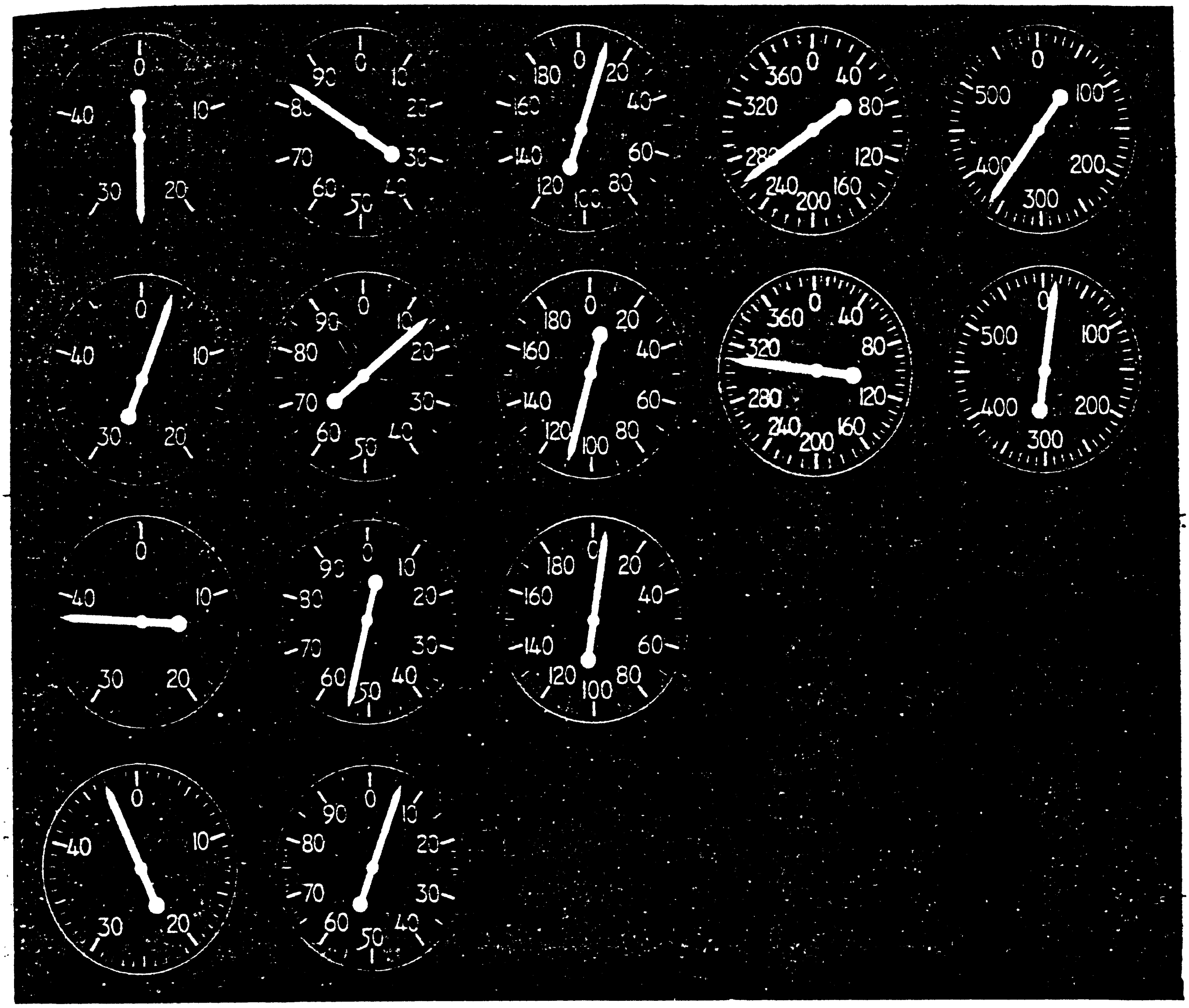


Figure 42, Dials Tested by Kappauf and Smith (1950)

Figures 43 (1.4 inch dials) and 44 (2.8 inch dials) show the frequency of reading errors of as a function of the "called interval," the unit to which readings are made. Notice that in both figures the y axis is logarithmic. In general, there were no differences in error data between the two dials sizes. In addition, the data show that tick marks should be provided down to the level to which a display is to be read (called). The pattern for larger errors (5 or more units) is similar.

Shown in Figure 45 are the data for reading times. Note the pattern is similar to the reading error data, though the graduation interval has less of an effect, if any.

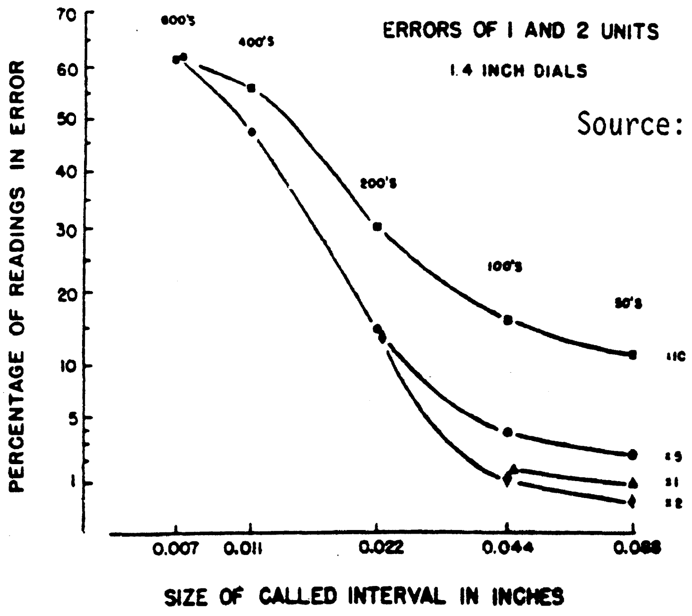
Kappauf (1951)

This study was concerned with examining plus ten errors on scales numbered in tens when small values were called, reversal errors at all values, the effect of the location of the numbers on reading performance (inside or outside of the scale), and display contrast direction (white on black versus black on white). In all, 40,400 readings were obtained from 46 subjects, a very large sample size.

Subjects, virtually all of whom had 20/30 vision or better, sat in a three-row section of bleacher seats 8, 10, or 12 feet from a screen. Dials, 12 per set, were mounted on plywood sheets covered by a black curtain. Panels were illuminated at a "daytime" level.

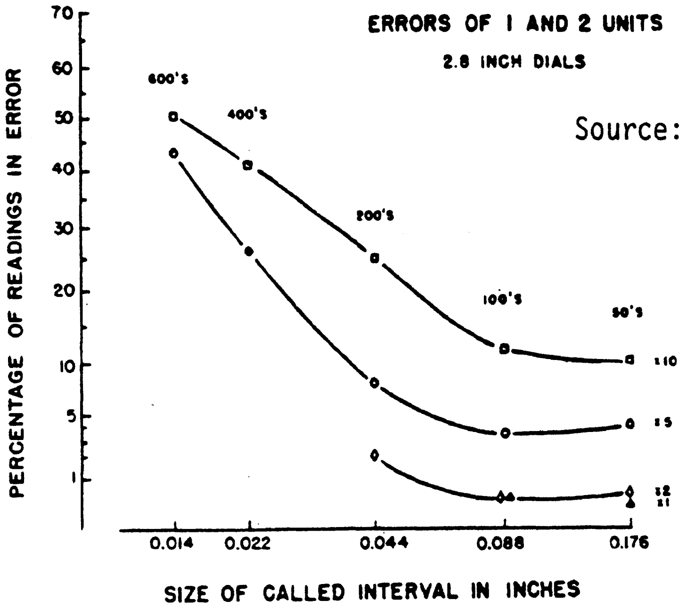
Shown in Figure 46 are 8 of the 16 designs examined. (The others were white on black.) Each dial was seen by each person in all of the 51 test positions. After viewing a dial array, the participant wrote down the values shown to the nearest unit and the elapsed time for reading that array.

With regard to errors, there were no differences between white on black and black on white dials. Likewise, differences in reading times were also not different (1.39 seconds for white characters, 1.38 for black). Kappauf also reported that the fewest errors occurred when the scale had a break and the zero value are near the bottom (98 vs. 134 errors). Most of the errors for other designs were reversals at zero.



Source: Kappauf (1951)

Figure 43, Frequency of Reading Errors for 1.4 Inch Dials



Source: Kappauf (1951)

Figure 44, Frequency of Reading Errors for 2.8 Inch Dials

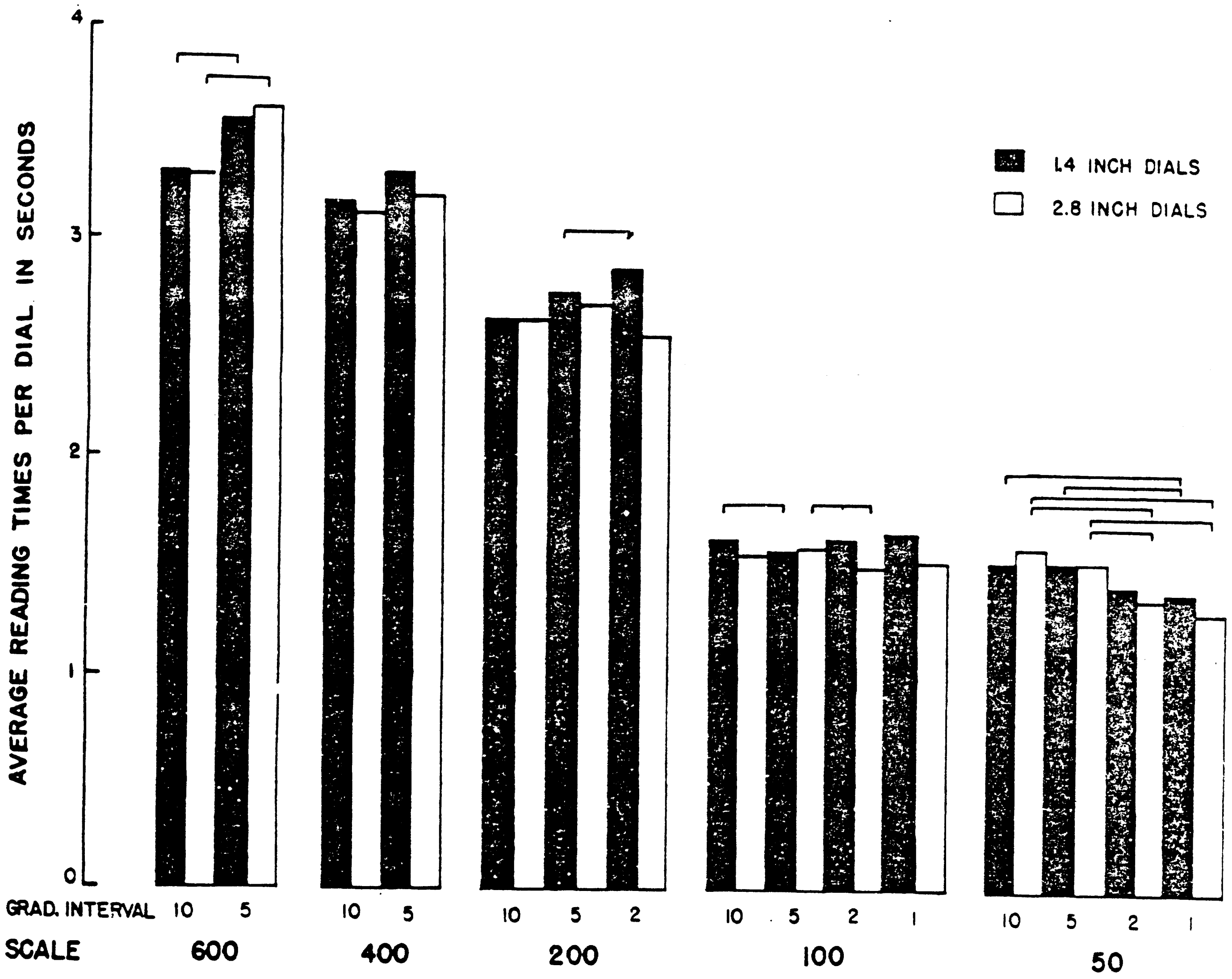


Figure 45, Reading Times for Dials in Kappauf and Smith (1950)

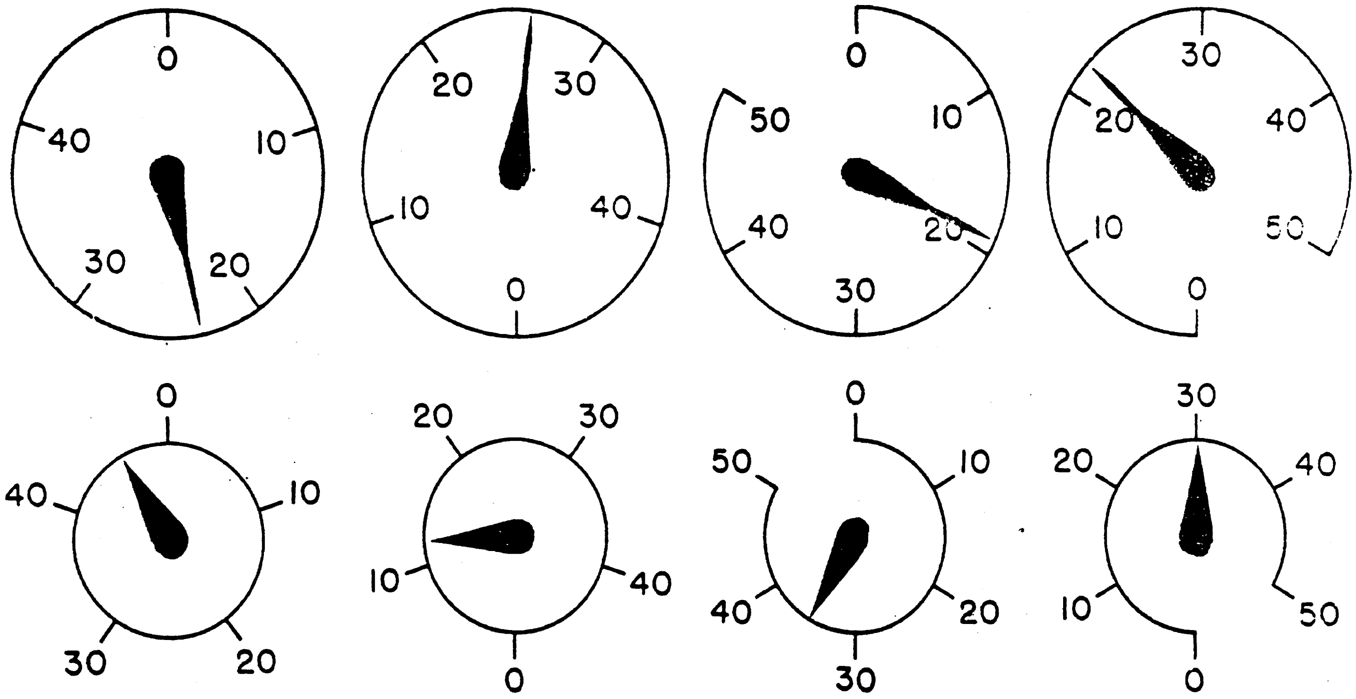


Figure 46, Dials Examined by Kappauf (1951)

Also, there were slightly fewer errors made when the numbers were outside the dial rather than inside (244 vs. 292), especially for high school students. On the other hand, responses to numbers outside the scale took slightly longer (1.52 vs 1.48 seconds). This suggests there is no difference due to number location with respect to the scale and pointer.

Carr and Garner (1952)

Several studies (e.g., Grether and Williams, 1949; Kappauf and Smith, 1950) examine the critical marked interval, a point beyond which increasing the distance between marks offers no improvement in performance. Often, however, the space available for a dial is limited. In that case the question is how many markings should be provided to maximize reading accuracy.

Nineteen people with 20/20 visual acuity or better were shown two markers in a lighted rectangle. The markers defined the ends of the scale. A pointer appeared between the markers. A total of 28 combinations were examined, 13 marked interval sizes (.5 to 25.0 mm) and 3 marker and pointer widths (.1, .2, .4 mm). The task was to estimate the pointer position to the nearest 1/100th of the marked interval. Each person made 101 judgments under each of the 28 conditions.

As noted in Figure 47 the relative error (in percent) was inversely proportional to the marked interval error (in mm). Beyond 15 mm additional markings did little to reduce the relative interval. However, the absolute error was proportional to the size of the marked interval (in mm). (See Figure 48.) Notice that both the marker and pointer width had little effect on performance.

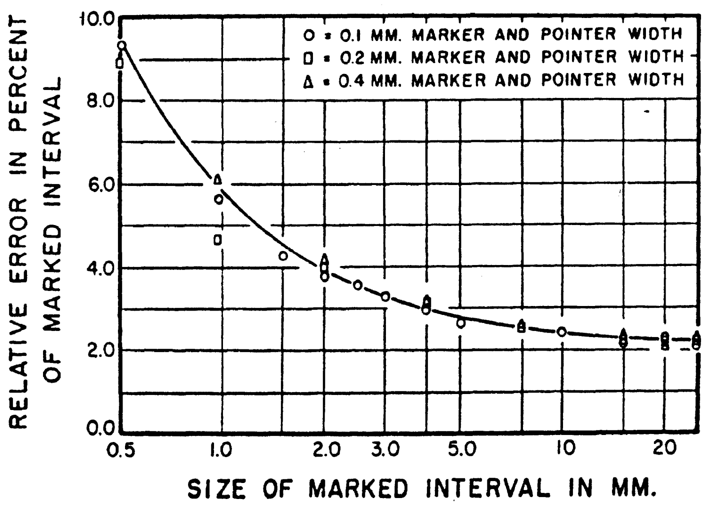


Figure 47, Relative Errors Reported by Carr and Garner (1952)

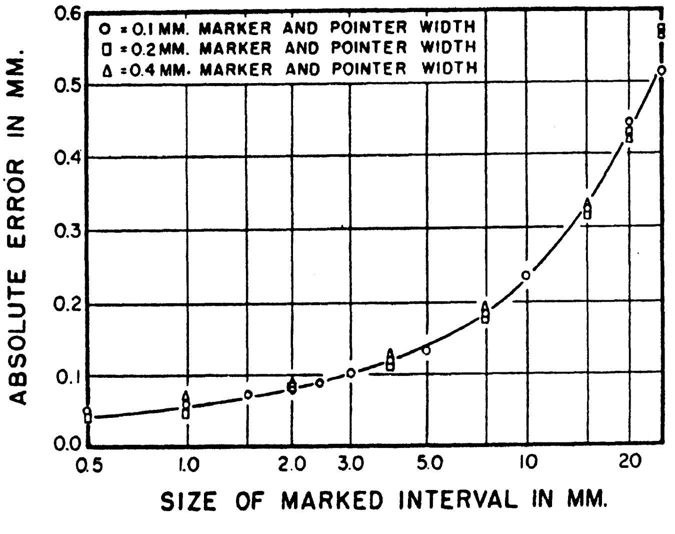


Figure 48, Absolute Errors Reported by Carr and Garner (1952)

White and Sauer (1954)

This study concerns the design of scale markings for reading under low illumination conditions resembling those found in cockpits at night. Eighteen male college students with 20/20 vision or better (determined at high brightness) participated. White on black scales printed on cards were viewed in a tachistoscope at 28 inches. (See Figure 49.) Scales displayed either 8 or 10 units. Four mark width and three interval sizes were examined. Large graduation marks were 1/2 inch long, intermediate marks were 1/4 inch long, and small graduation marks were 1/8 inch long.

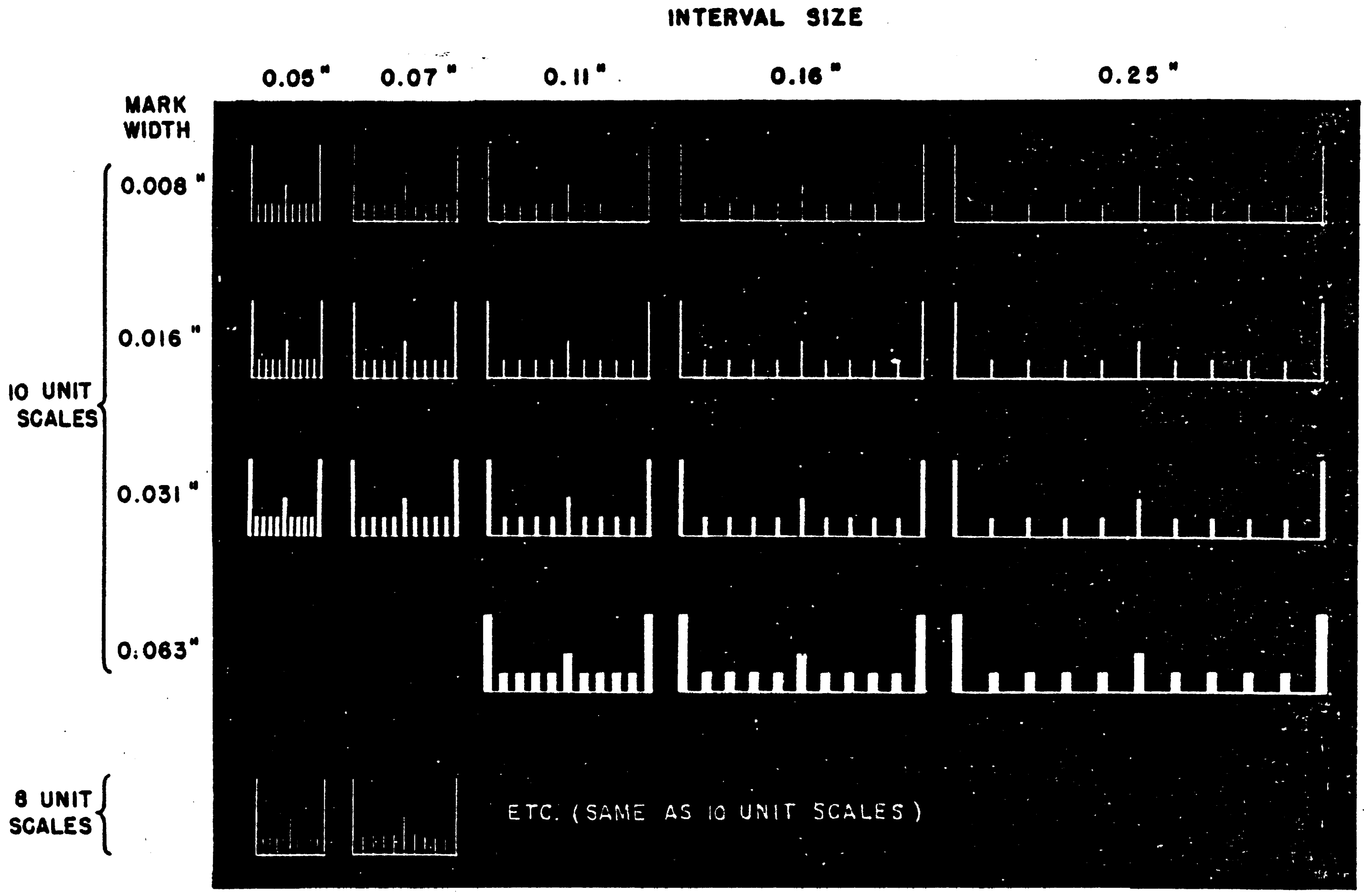


Figure 49, Scales Examined by White and Sauer (1954)

After 10 practice trials, each participant viewed each of the 36 scales twice at each of 3 luminance levels. They were .1 ft-L (representative of night lighting of aircraft instruments), .01 ft-L, and .002 ft-L. Participants pressed a button to view each scale and released it when a clear reading of the exposed scale was obtained. Scale readings were reported aloud.

There were statistically significant differences due to mark width, interval size, and luminance. There were also significant interactions between width and interval, and width and brightness. The scale length had no effect on performance.

Figure 50 shows the reading time and error data. Notice that both time and errors increase drastically for the .002 ft-L condition when the marker width is less than .031 inches or the interval size is less than .11 inches. For higher levels of luminance performance suffers most when the interval size is less than .11 inches. In that instance mark width is of secondary importance. It appears that the recommendations in the Military Handbook (U.S. Department of Defense, 1981a) and 1472 standard (U.S. Department of Defense, 1981b) are based on this experiment.

Churchill (1956)

Ten employees of a Canadian defense laboratory viewed scales 28 inches away. Scales were black on white. Six scale intervals and six pointer clearances were examined. The illumination level was 180 ft-c.

Participants read the scales as rapidly and accurately as they could. In the first half of the experiment the participant pressed a button to close the shutter. In the second half, the exposure duration was fixed at 300 ms. Each subject saw 18 random settings for each of the 36 pointer-scale interval combinations.

In an ANOVA of the transformed error data, both the effects of pointer clearance and scale interval length were significant. Figure 51 shows the performance data. Notice that the smallest number of errors is for a scale interval length of 1.5 inches, slightly larger than that recommended by Grether. Errors were minimized when pointer clearance was .25 inches, though the differences in performance between zero clearance and .50 inches were small. Readers should bear in mind that this experiment refers to instances were scale marks are spartan and interpolation is always required, not necessarily the typical case in automotive contexts.

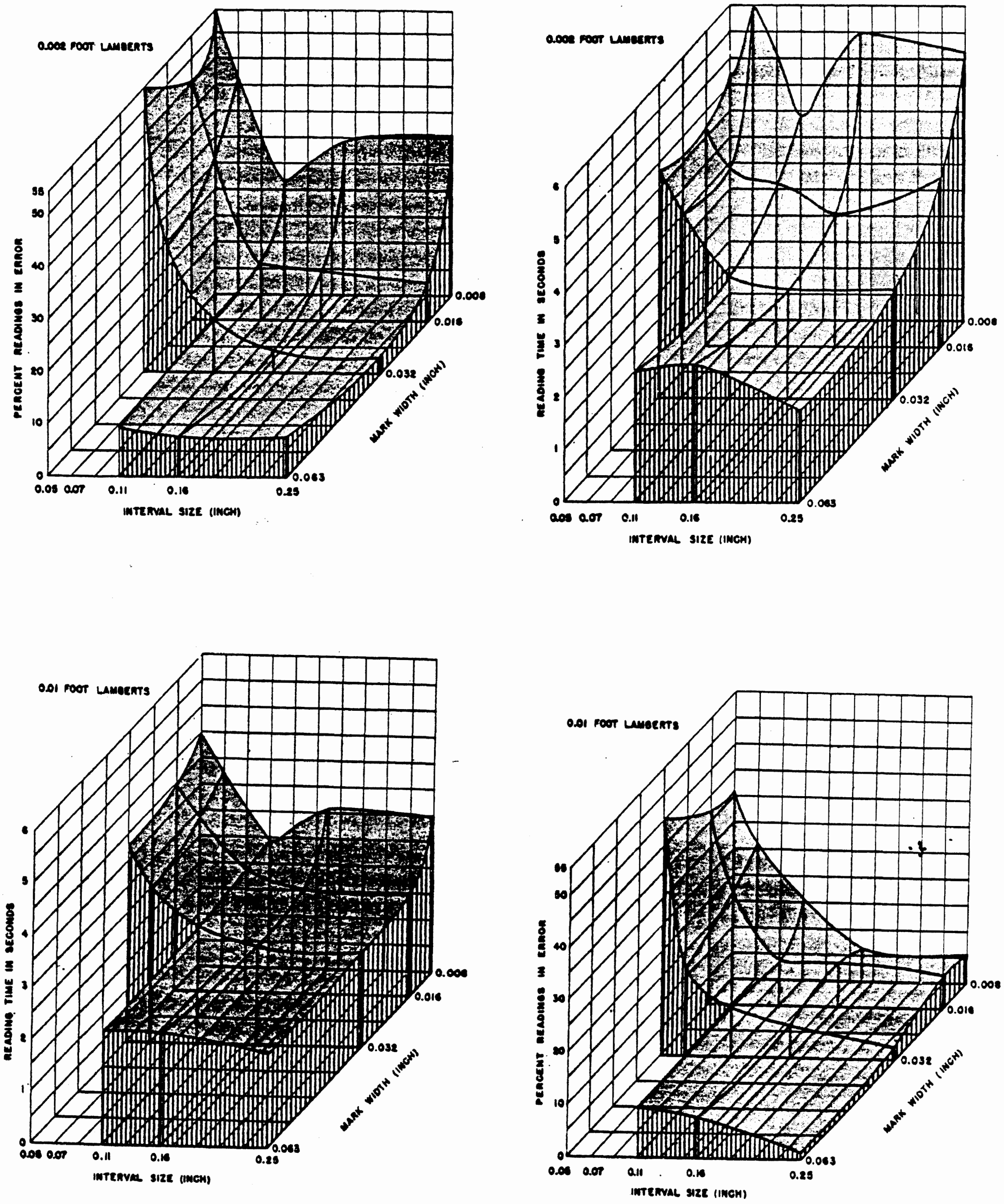
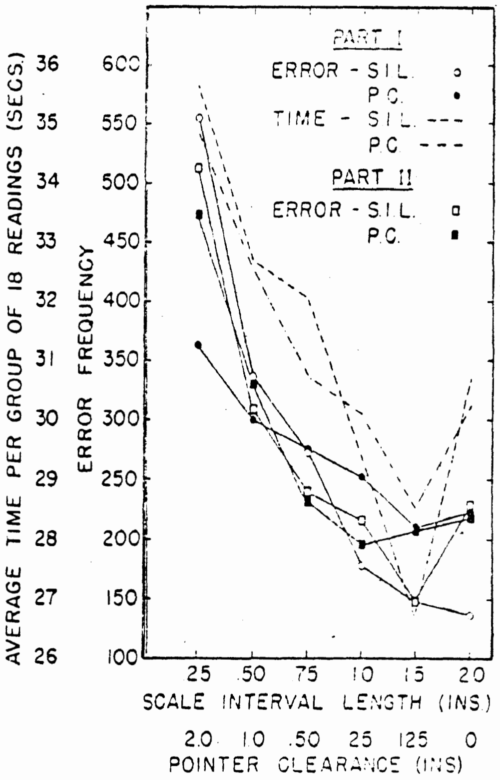


Figure 50, Results from White and Sauer (1954)



Source: Churchill (1986)

Figure 51, Effect of Scale Interval Length and Pointer Clearance on Performance

Cohen and Dinnerstein (1958)

It is well known that reading nonlinear scales is difficult, but there is limited quantitative data on the extent to which it is a problem. Sixty-four college students viewed dials through a tachistoscope. Just before the dial was shown each participant was told the expected reading. Their task was to say if the value shown was within 1 graduation mark, a check reading task. They held down a button to control the exposure duration.

Dials were black on white with a luminance of 9 ft-l. Four dial designs (shown in Figure 52) were examined. Four hundred copies of each dial were printed on cards and the pointers were drawn in by hand. For each of the 4 dials 4 distributions of pointer settings (corresponding to the 4 scale types) were examined. Hence for each dial design there was one condition where the distribution of pointer settings was uniform.

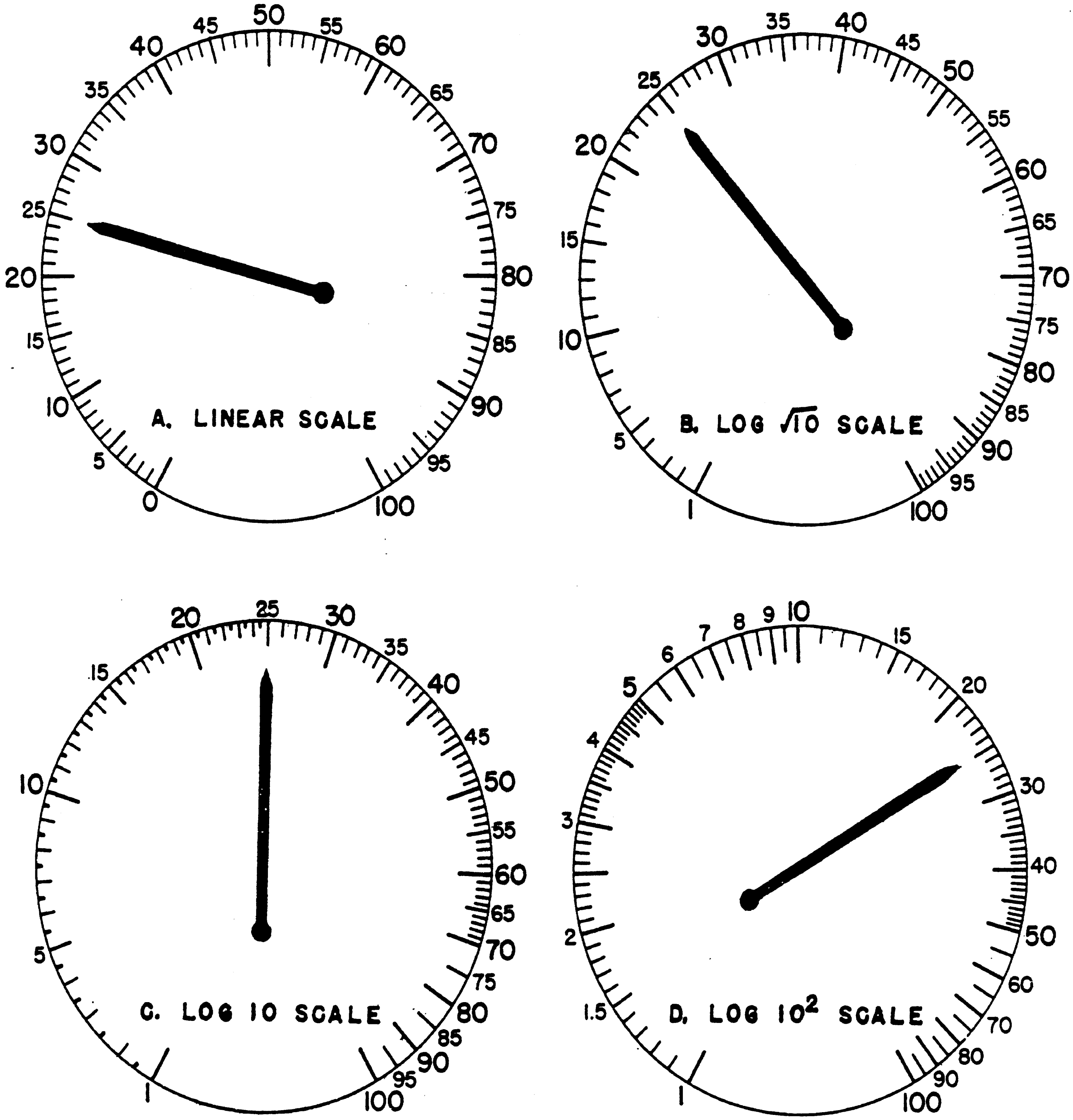


Figure 52, Dials examined by Cohen and Dinnerstein (1958)

While linear scales were read more accurately in general, there were no overall statistically significant differences between scales with regard to errors. There was, however, a small difference in the time required to read scales, with the linear and log square root 10 scale being read in less time than the others. While there was a good correlation between the expected rank order of errors (linear distributions being accurately read on linear scales, etc.), the interaction between scale type and distribution was not statistically significant. (See Table 23.) In part, this was because of how the experiment was designed--each subject saw only 1 of the 16 scale-distribution combinations and the variability due to subjects was large. Apparently in some cases linearizing the distribution of readings increases the dispersion of the readings which can make a display more difficult to read. This does not necessarily mean that nonlinear scales should be used, as the loss of accuracy in certain portions of the scale range can cause problems. However, if nonuniform accuracy is required, for example if there is need for greater accuracy at the low end of the scale, then a log scale may be beneficial.

Table 23, Error Percentages from Cohen and Dinnerstein (1958)

	Dial			
Distribution	log 10 sq	log 10	log sqrt 10	linear
log 10 squared	3.4	5.4	8.3	1.7
log 10	3.8	4.6	7.5	5.4
log sqrt 10	4.2	7.9	8.8	4.2
linear	8.8	5.4	5.0	3.8

Murrell, Laurie, and McCarthy (1958)

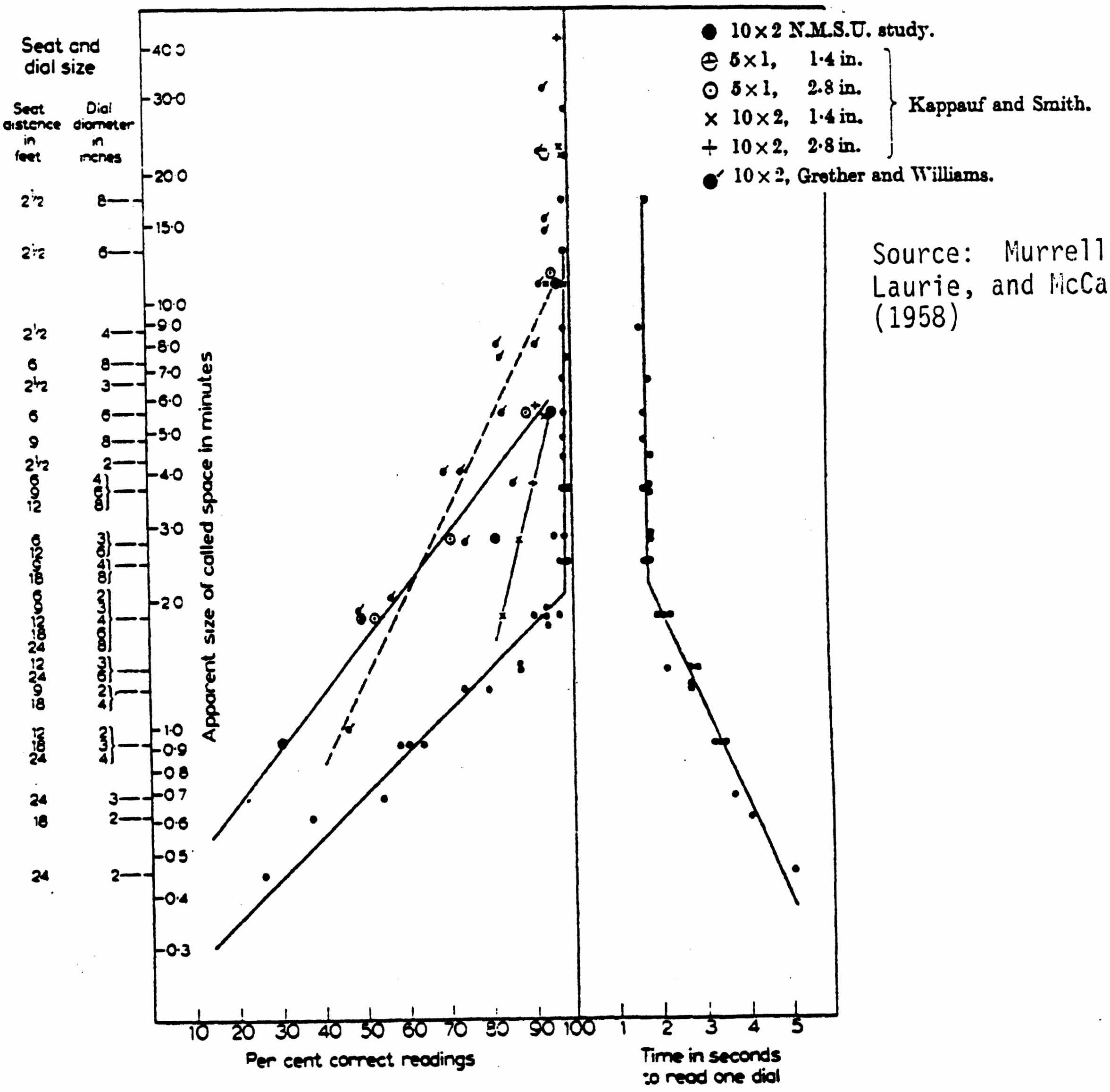
This study examined the relationship between reading distances and optimum dial sizes. As noted earlier, errors decrease rapidly until the marked scale interval reach about .25 inches, after which errors decrease at a slower rate. There is debate as to whether the minimum point is .5, 1. or 1.5 inches.

In this experiment groups of 12 dials were mounted on a board and read by 6 petty officers at their own pace. They also timed themselves. They all had 20/40 vision or better.

Five dials sizes were used, 2, 3, 4, 6, and 8 inches in diameter. The actual scale diameter was 5/6 the blank dial diameter. Dials were 270 degree 200-unit scales numbered in 20's with tick marks at 10's. Each dial was read from 6 distances (2-1/2 to 24 feet).

Shown in Figure 53 are the results from this and other studies. Number pairs (e.g., 10 x 2) refer to the scale interval (tick marks) and the called interval, the accuracy to which dials were read. The N.M.S.U. study (Naval Motion Study Unit) is the

Murrell, Laurie, and McCarthy paper. Notice that the number of errors begins to increase when the called space (how accurately the scale is to be read) is less than 13 minutes of arc, though there are no consequences in terms of time until the spacing is 2 minutes of arc. (Elsewhere (Murrell, 1969) the 2 minute value is treated as the critical number.) For an instrument panel viewed at 28 inches a called space of 13 minutes of arc is equivalent to just over 1/10 of an inch on the dial. Readers should bear in mind that these data are for highly practiced subjects who have 20/20 visual acuity or better. For drivers, many of whom do not have 20/20 acuity, nor the practice that Murrell's participants had (2000 readings) larger values should be used. Details of the calculation procedure are described in Murrell's 1969 book and it is summarized later in this section.



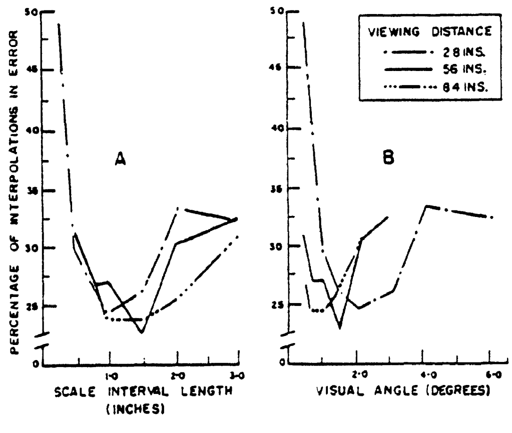
Source: Murrell, Laurie, and McCarthy (1958)

Figure 53, Interval Spacing Data from Several Studies

Churchill (1959)

This study extends the results of the previous one to other viewing distances. Seven scales were viewed by 24 Canadian defense laboratory personnel at 28, 56, and 84 inches, exposed for 500 ms. They saw 18 settings at each distance. The luminance level was 120 ft-L. Their task was to read each scale to the nearest unit. (The scale was 0 to 10 with only 0 and 10 marked).

As shown in Figure 54, errors were at a minimum when the scale interval was 1 to 1.5 inches and changed only slightly with changes in viewing distance. This is evidence that the law of constant visual angle does not always hold.



Source: Churchill (1959)

Figure 54, Effect of Viewing Distance on Scale Reading Errors

In a second experiment black on white scales were viewed by five laboratory personnel. Components of the intervals (line thicknesses, pointer dimensions, numeral size, etc.) were scaled so their visual angle was constant. In the first experiment their size was fixed, so their visual angle varied with the test condition. Here again, three viewing distances were used. As noted in Table 24, errors were minimized when the scale interval was 1.5 inches. Hence these results suggest that actual size is more important than visual angle.

Table 24, Errors as a Function of Visual Angle in Second Experiment

Viewing Distance (inches)	Scale Interval Length (inches)				
	0.5	1.0	1.5	3.0	4.5
28	25.7		6.7		
56	35.7	14.8	8.5	19.0	
84	46.2		9.4		23.7

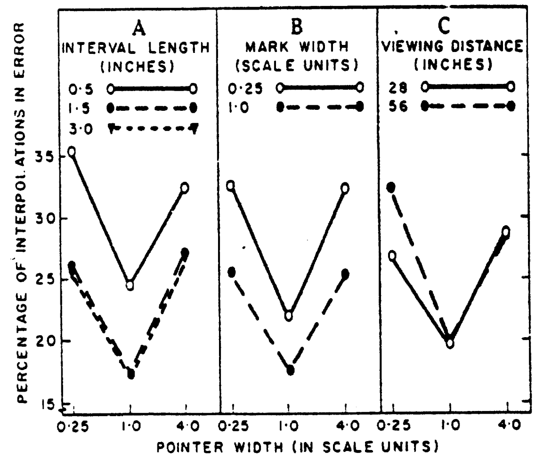
Source: Churchill (1959)

Churchill (1960)

One weakness with the previous studies was that the relative size of the pointer to the interval varied. For example, in Grether and Williams (1949) the pointer width was .094 inches for all interval lengths (.32 to 1.31 inches or a range of 1:0.7 to 1:31.3 scale units).

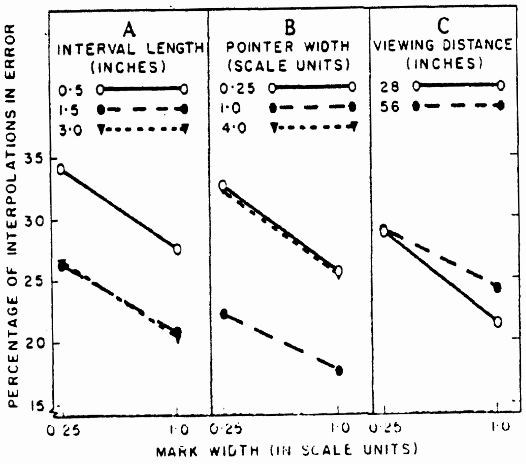
Displays consisting of three interval lengths (0.5, 1.5, 3.0 inches), three pointer widths (.25, 1.0, 4.0 scale units), and two scale mark units (.25 and 1.0 scale units) were shown at 28 and 56 inches to 10 women. The display luminance was 2.0 ft-L for all conditions. Each dial was exposed for 250 milliseconds.

As shown in Figures 55 and 56, errors were minimized when the mark and pointer width were 1 scale unit wide (for a scale from 0 to 10 without other marks). This was true in general for both viewing distances.



Source: Churchill (1960)

Figure 55, Pointer Width Versus Other Factors



Source:
Churchill (1960)

Figure 56, Marker Width Versus Other Factors

Kelso (1965)

This study concerned the design of a moving scale (tape) display. See Figure 57 for examples. Kelso varied the spacing of the marked intervals (1-3/8 to 2-3/8 inches), the number of graduation marks between them (0, 1, 3, 4, or 9), and the scale orientation (vertical or horizontal). A total of 150 Air Force officers, all with 20/20 vision were shown subsets of the slides of the scales. After the participant had seen the display, he pressed a button to close the shutter and then said the value aloud.

The scale factor, and especially graduation marks, had statistically significant effects on both reading time and errors. Horizontal scales were read more rapidly but not more accurately than vertical ones. The best performance was for a marked interval spacing of 1-7/8 inches. There were large differences due to the number of graduation marks with error rates declining dramatically as scale marks were added from 0 to 9 (27, 21, 20, 10, and 9% respectively). For the response time data the ranking of marks from best to worst was 9, 1, 3, 4, and 0.

Chapanis and Scarpa (1967)

This is another study that examined the readability of dials at different distances when the angle subtended was held constant. The common notion is that beyond 2 meters one is only concerned with visual angle. This study supported that idea.

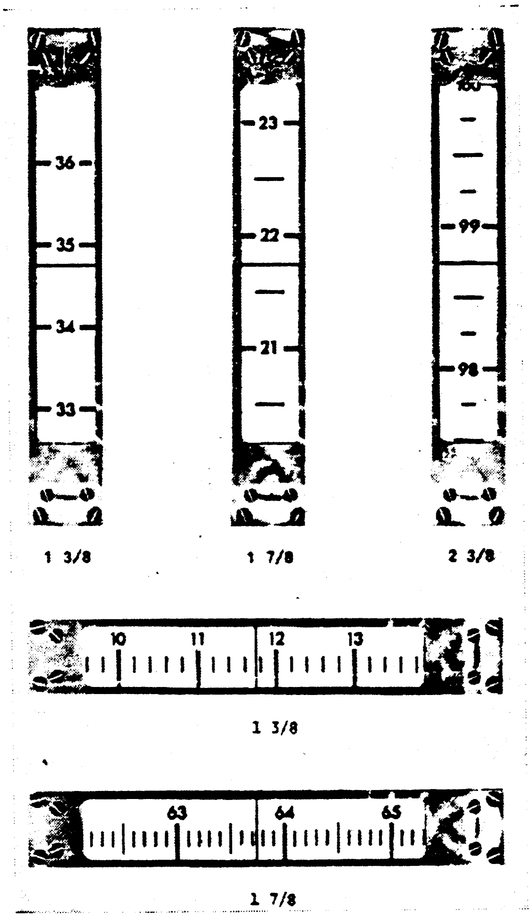


Figure 57, Moving Tape Displays Examined by Kelso (1965)

Five dials were read 20 times by 20 people with normal visual acuity. Verbal response times were recorded. Each dial was seen at only one viewing distance (14-224 inches) but was sized so it occupied the same visual angle.

There were significant differences among dials with dials at greater distances being read more quickly. Dials at greater distances were also more likely to be read accurately. Finally, subjects said that the dials at a greater distance were easier to read. Hence, this study supports the conclusion that the actual size of dial markings is what matters most, not their visual angle. This suggests that recommendations for scale markings for viewing distances obtained by simple adjustments for visual angle should be used with care. This does not create a problem for automotive applications. Most of the studies described in this report involved viewing distances of approximately 28 inches, typically what one finds in cars and trucks.

Murrell (1969)

This book, referred to earlier, provides a procedure for calculating required dial sizes given the accuracy to which they must be read and the viewing distance. The procedure is based primarily on his 1958 data described earlier (Murrell, Laurie, and McCarthy, 1958). The critical assumption is that the optimum called scale division (how accurately a scale is read independent of how it is marked) should subtend at least two minutes of arc. Again, this number is based on data with 20/20 visual acuity (which only some drivers have) for highly practiced subjects.

Murrell (1969) argues that this figure should be increased by only 1/5 to allow for observers with less good eyesight and where conditions are less than ideal. For example, the author would argue for doubling it to accommodate observers with 20/40 visual acuity, and at least doubling it again to account for less than ideal viewing conditions. In fact, based on Murrell's 1958 data, using 13 minutes of arc seems more appropriate. The 2 minute rule leads to mark spacings of .016 inches, which seems unreasonably small. Using 13 minutes of arc, the spacing is .105 inches.

Using Murrell's 2 minute of arc value,

$C/D >= \tan(2')=5.8177E-04$

where: C=Called interval
(smallest interval to which is read)
D=Viewing distance

or more generally

$C=D \times \tan(a)$

where: a is the minimum acceptable angle for the called interval, 13 minutes being recommended by the author

If a scale has n discrete points on it (e.g., a scale from 0 to 100 read to the nearest unit would have 101 discrete points, then the scale base length (B) is:

$B=n \times C= n \times D \times \tan(a)$

For circular scales, the base length (circumference) is equal to 2 x pi x diameter. If a 60 degree gap is left between 0 and the maximum, then the base length is 5/6 x 2 x pi x diameter. (Note: (360-60)/360=5/6.) Dividing through, the dial (DD) diameter is

$DD = n \times D \times \tan(a) \times 3 / (5 \times \pi)$

On pages 194-196 Murrell provides expression for calculating recommended dial sizes. Some of the expressions given elsewhere for dial sizes may be in error and that could also be true here. Readers interested in such calculations should rely upon their knowledge of trigonometry and the information above for making such calculations. Clearly a major assumption is the minimum visual angle for the called interval, and depending on how the data in the literature are interpreted, recommendations for dial size can therefore vary quite widely.

Summary

A large number of studies have been conducted to examine scale marking, probably more than those which examine any other subject related to scales except pointer alignment. These studies show the following:

1. When reading groups of scales, it is important that the marking scheme across scales be consistent. Inconsistency can double the time to read a set of scales.

2. Scales marked with values less than 1 are hard to read. Whenever possible multipliers should be provided so numbers appearing on scales exceed 1.

3. Scales should be marked in numbers that are even multiples of 10. Use of other schemes can triple the error rate.

4. Arranging minor marks in a staircase fashion does not significantly reduce reversal errors, though it does help some.

5. If lighting is adequate to read a gauge, going much beyond adequate has a minimal effect on performance.

6. Black on white dials are just as easy to read as white on black.

7. At standard instrument panel viewing distances (about 28 inches), dial diameter has a minor affect on reading time if a single dial is in a known position on which one is fixating.

8. It doesn't seem to matter if numbers for dials are on the inside or outside of the scale.

9. Dials without breaks (between 0 and the maximum) are difficult to read.

10. Scale marks should be at least .031 inches wide. The literature suggests that mark width (and for that matter pointer width) are not critical factors, and variations of them will have only small effects on performance.

11. There is debate in the literature concerning the spacing of called interval marks. Based on one study the range of

recommendations is 2 to 13 degrees visual angle. (At 28 inch viewing distances these numbers are equivalent to .016 to .105 inches separation). In another study the recommendation is .11 inches. Just over a tenth of an inch seems to make sense.

12. There is still controversy as to how far apart marked intervals should be on a scale. Recommendations range from .5 to 1.5 inches.

13. Transforming a scale (e.g., using a log) to linearize the distribution has only a small benefit on practiced reading performance.

14. Scales with a perimeter (baseline) are much easier to read than those without. Baselines may be one of the most important elements of a dial.

15. Recommendations for display design parameters are often not a simple function of visual angle. Hence, it is sometimes risky to extend the recommendations obtained at one viewing to another using the ratio of the viewing distance. This is particularly true, say for comparison of 28 inches versus 10 feet.

Finally, while the literature offers many specific recommendations about scale design, data on tradeoffs is weak. (The subject of tradeoffs is covered in a later section.) In brief, the automotive designer is often given a limited space into which a display must fit. Quite often, if the human factors recommendations for marked interval spacing, major and minor mark size, and so forth are followed, the displays will not fit. It is difficult for the designer to decide what to do in such circumstances.

HOW SHOULD ZONES BE MARKED?

The number of studies that have considered the effectiveness of alternative ways of marking scale zones is limited. The most important study for automotive applications, Green (1984), is described in detail in this section.

Kurke (1956)

There are only a few studies in the literature that deal with the coding of regions on displays. Kurke (1956) describes a dial designed so that when the pointer position indicates caution or danger condition, a high-contrast wedge appears on the dial face that is not present when the pointer is in the "safe and normal" limits. (See Figure 58.) Kurke reports that favorable comments were received from helicopter pilots during informal testing.

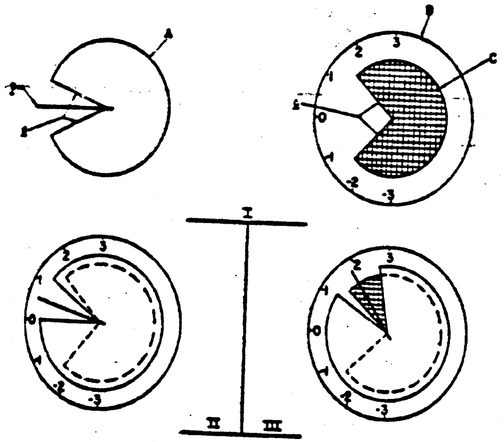


Figure 58. Dial Principles of Operation (Kurke, 1956)

For a formal test, 4 decks of 50 index cards were drawn. Figure 59 shows examples from 3 decks. The fourth deck had numbers (in black and red) and circles (either filled or not filled) on them. Thirty-three men sorted each deck into two piles, turning the cards over one at a time. Cards were sorted by condition (safe/normal versus redline), and the numbered deck by color, whether the circles were filled in, and so forth. Both time and errors were recorded.

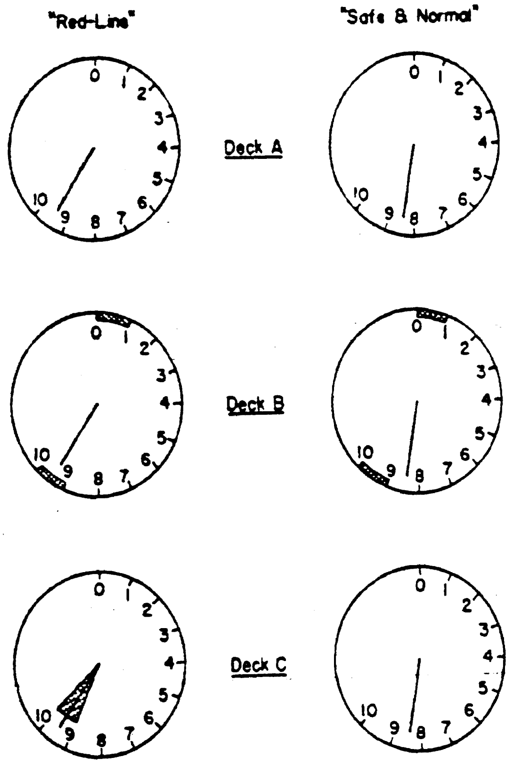


Figure 59. Dial Designs Examined by Kurke (1956)

Out of 1650 trials only 1 error was made with the exposed area dial (deck C), while 18 errors were made with the conventional red line design (deck B) and 39 when no markings were provided (deck A). Sort times for those decks were 52.9, 60.6, and 73.1 seconds respectively. In general, differences between the display designs were statistically significant, demonstrating the benefits of the experimental markings on performance, and zone coding in general.

Sabeh, Jorve, and Vanderplas (1958)

This study was carried out to investigate which shapes should be used as marking bands on dials. It was conducted because color coding was ineffective when red lighting was used. Based on a preliminary survey (involving 1500 and shapes 79 subjects), the seven marking schemes shown in Figure 60 were chosen for further study. Two versions of a questionnaire were developed, one with numbers shown on the dials (as in the Figure), and one without. The surveys were given to 70 pilots and 70 college students. They wrote down the category of warning that each marking best represented.

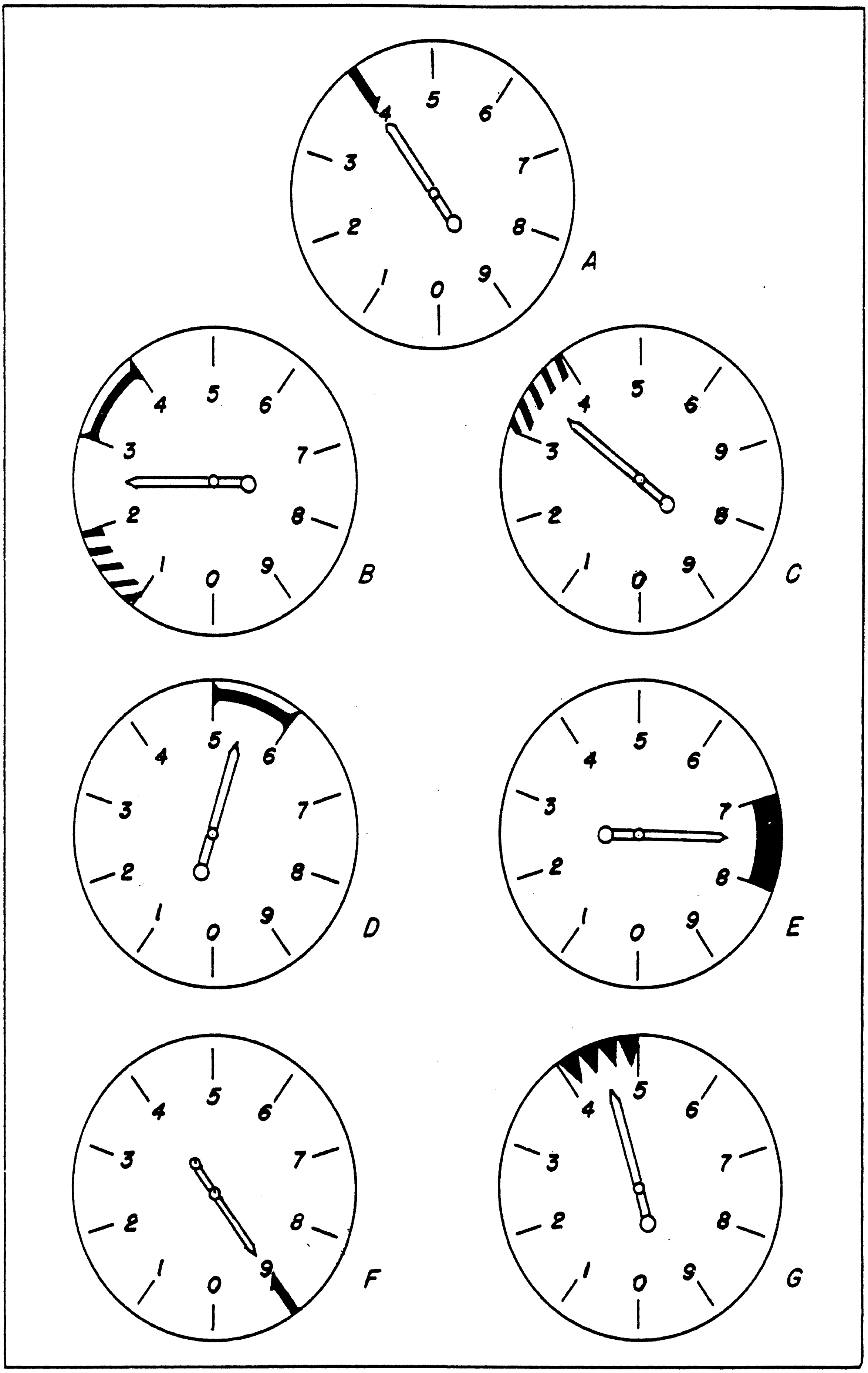


Figure 60. Markings Examined by Sabeh et al (1958)

Table 25 contains the significant pairings and the Table 26 the tallied preferences. Note that while the data show consistent trends, there were only a few cases were a particular pairing was preferred by a majority of the participants, and in no cases was there 100% agreement. That does not mean, however, that this effort was wasted. While some learning is required here, the amount required is much less than had the pairing been arbitrary. Further, while the application of the survey was towards aircraft engine displays, these data should be applicable to automotive displays as well.

Table 25. Significant Associations (Sabeh et al. 1958)

Shape	Pilots	Students
A	Danger-lower limit	Danger *
B	Undesirable	Caution
C	Caution	Caution **
D	Mixture-lean	Mixture-lean **
E	Mixture-rich	Mixture-rich
F	Danger-upper limit	Danger ***
G	Dangerous vibration	Dangerous vibration

* lower limit when numbers provided, upper limit without
** without numbers only
*** upper limit with numbers, lower limit without

Wokoun and Chaiken (1959)

This report does not document any experimental evidence but as its title indicates is: "A Guide to Color Banding for Indicators." It recommends that displays should be coded to follow the commonly accepted stereotypes, though no empiric evidence is provided to support their recommendations.

1) Red (Munsell 5R 4.5/14) should be used to indicate "a dangerous condition requiring immediate corrective action."

2) Green (Munsell 5G 6.1/11) indicates the "normal" or "acceptable" operating range.

3) Blue (Munsell 8PB 3.1/12) is used with green to indicate a secondary desired range.

4) Brown (1YR 4.1/4) is to be used for a tertiary operating range, but only if "absolutely necessary." The author does not know of any circumstances where brown has been used to color code scales.

Table 26, Frequency of Association of Zone Markings with Categories

STUDENT GROUPS

	With Numbers							Without Numbers						
Shape--	A	B	C	D	E	F	G	A	B	C	D	E	F	G
Category--														
Danger-Lower Lim.	(12)*	5	9	1	0	7	1	6	6	1	1	2	(14)	5
Caution	1	(11)	4	6	10	0	3	1	(14)	(11)	5	1	1	2
Undesirable	7	5	5	5	9	3	1	3	7	8	4	6	3	4
Mixture-Lean	3	6	7	9	3	3	4	6	1	7	(12)	3	4	2
Mixture-Rich	5	2	8	9	(11)	0	0	2	2	5	6	(15)	3	2
Danger-Upp Lim.	6	2	0	4	1	(20)	2	(15)	2	1	5	3	8	1
Dangerous Vib.	1	4	2	1	1	2	(24)	2	3	2	2	5	2	(19)

PILOT GROUPS

	With Numbers							Without Numbers						
	A	B	C	D	E	F	G	A	B	C	D	E	F	G
Danger-Lower Lim	(27)	3	2	0	0	2	1	(20)	1	1	0	2	10	1
Caution	0	4	(20)	3	4	1	3	0	7	(23)	1	2	0	2
Undesirable	2	(14)	8	6	2	0	3	2	(16)	6	3	5	0	3
Mixture-Lean	1	9	1	(18)	3	3	0	2	7	2	(19)	3	1	1
Mixture-Rich	2	2	1	5	(24)	0	1	1	2	0	10	(19)	1	2
Danger-Upper Lim.	1	0	0	2	1	(28)	1	10	1	1	1	0	(21)	1
Dangerous Vib.	2	3	2	1	1	0	(25)	0	1	2	1	4	2	(25)

*Parentheses denote those frequencies found significant at the 1% level

Source: Sabeh et al. (1958)

5) Surprisingly, Wokoun and Chaiken recommend that amber not be used but that regions associated with neither normal operation or danger be left unmarked (white). This recommendation is quite different from what is common automotive practice.

Finally, they recommend that the width of color bands be equal to the length of the minor scale marks (tick marks).

Green (1984)

Probably the most important study concerning the marking of gauges was carried out by the author (Green, 1984). In fact, it is the only study in the literature that comprehensively examines scale marking and selection for automotive applications. The study examined 6 issues:

1. Should fuel displays be analog or numeric?
2. Should the same display format be used for all engine displays?
3. Does adding color coding lead to a significant reduction in gauge reading errors?
4. How should scales for engine gauges be labeled?
5. Which display for fuel level is best understood?
6. In general, what do drivers know about cars and what does that suggest about display design?

Sixty-six drivers (ages 18-78) carried out 4 tasks each. First, they were asked a series of questions about their cars (fuel capacity, normal engine temperature, etc.). These questions were asked to help explain why particular displays might not be understood.

Second, drivers sat in a mockup of a car and were shown slides of 72 instrument panels (from a set of 144). For each slide the driver stated if each display was high, low, ok, or if they were unsure. Further, they were given a list of potential actions and asked what they would do on a hypothetical trip if a display was not ok (ignore it, speed up, slow down, stop at the next service station, etc.).

Each instrument cluster contained a speedometer, three engine displays from a set of five (engine temperature, oil pressure, oil level, electrical system voltage, electrical system current), a fuel display, and several warning lights. Engine displays could be either numeric or moving pointer displays. For moving pointer displays, scales could be either labelled with numbers (e.g., 140...260), letters (C,H), or words (cold, hot), the pointer could be white or color coded (red/green), and a range mark ("ok") may or may not have been provided. Further, for moving pointer displays, the value could be indicated by illuminating one segment of a simulated electronic display (cursor design) or all points up to it (fill design). In a few cases both English and metric units were examined.

For the fuel gauge, 23 different designs were examined. (See Figure 61.) They varied in terms of whether the display was a moving pointer or numeric format, and, for each type, what the referrence value was.

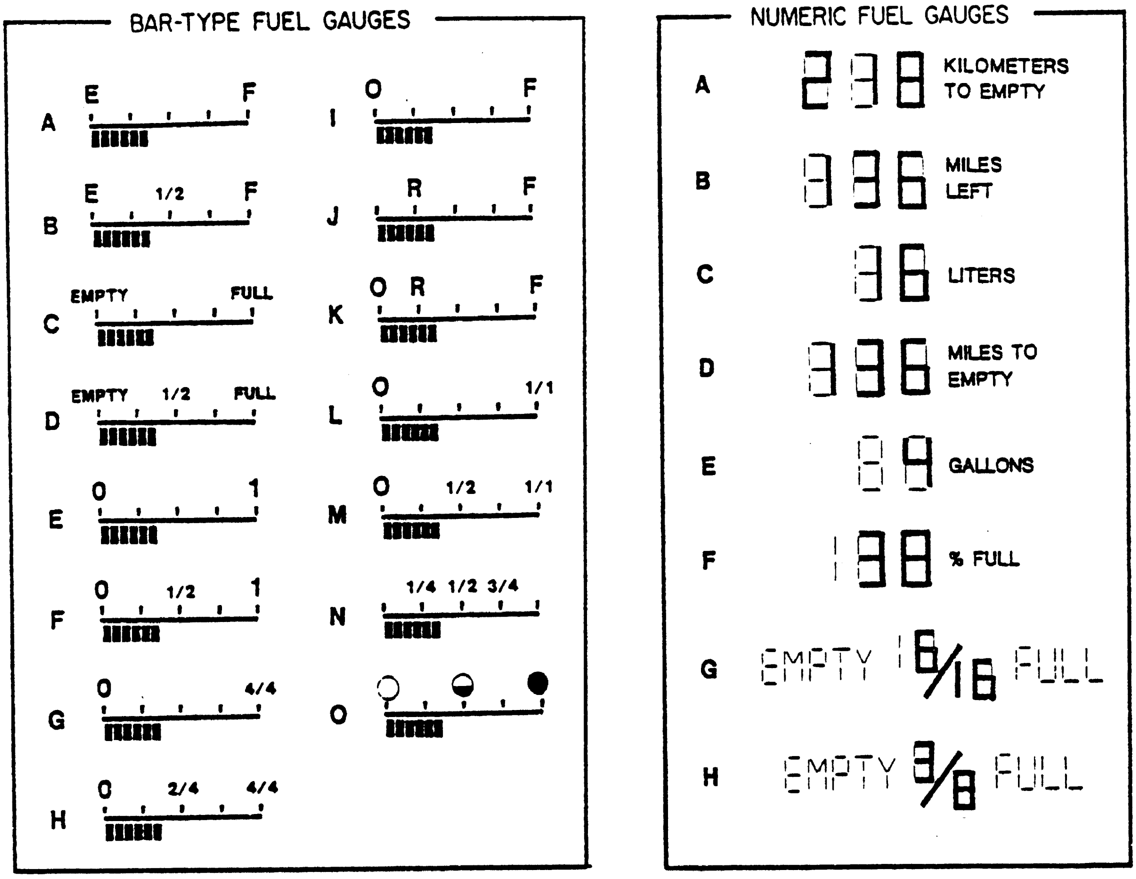
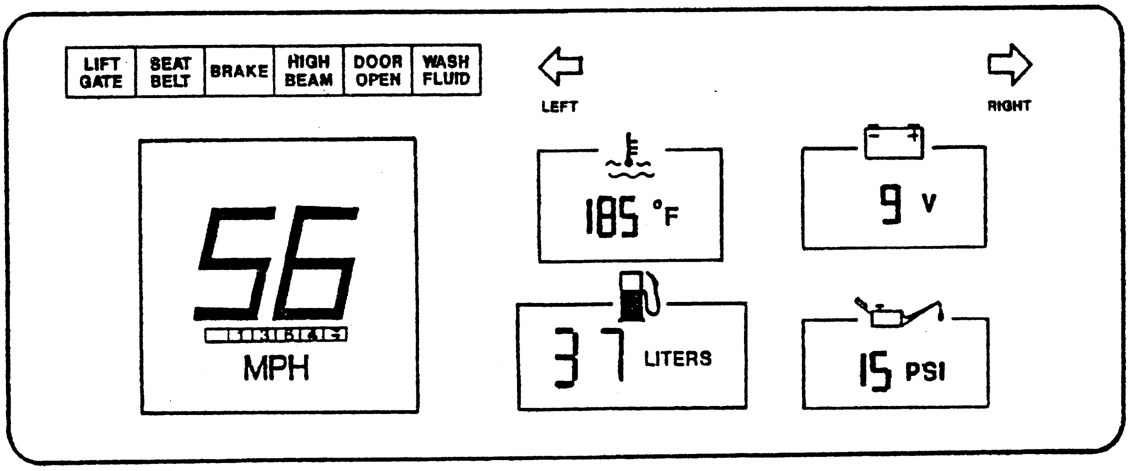
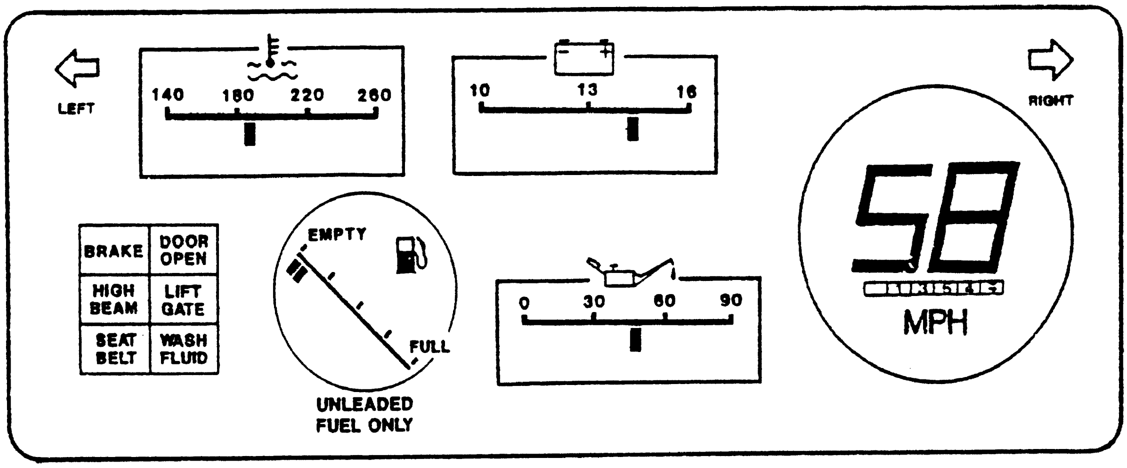


Figure 61. Fuel Gauge Designs Evaluated by Green (1984)

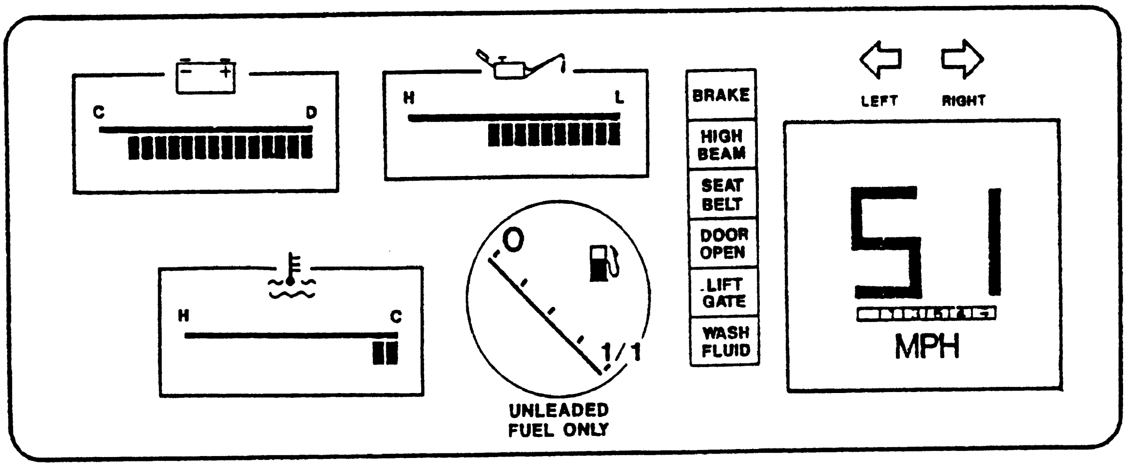
Examples of some of the clusters examined are shown in Figure 62.



Cluster with all numeric gauges - Class A1.

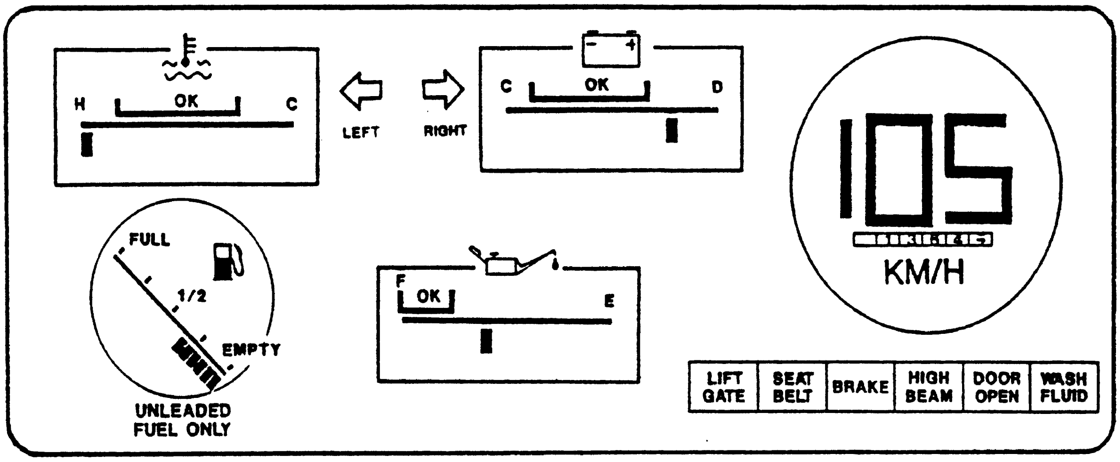


Cluster with plain bar-type displays - Class B1.

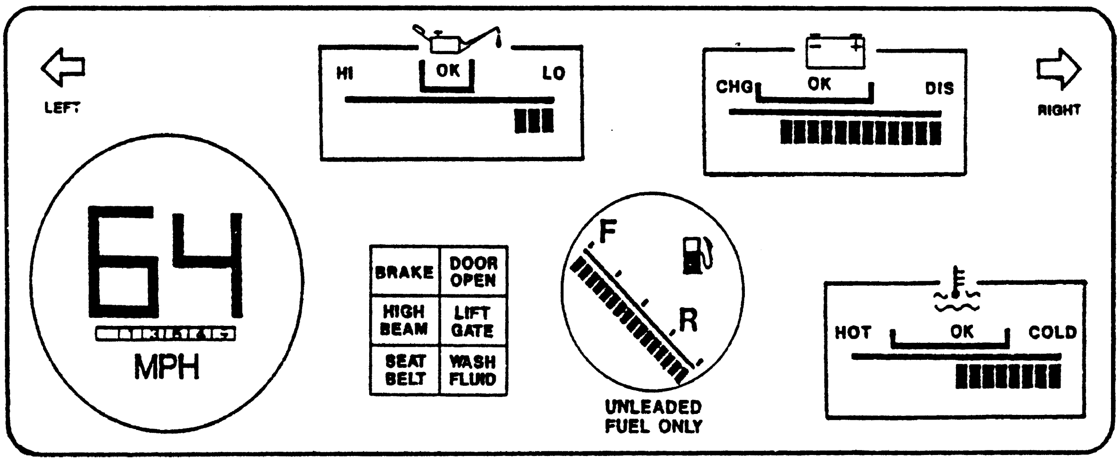


Cluster with plain bar-type displays - Class B2.

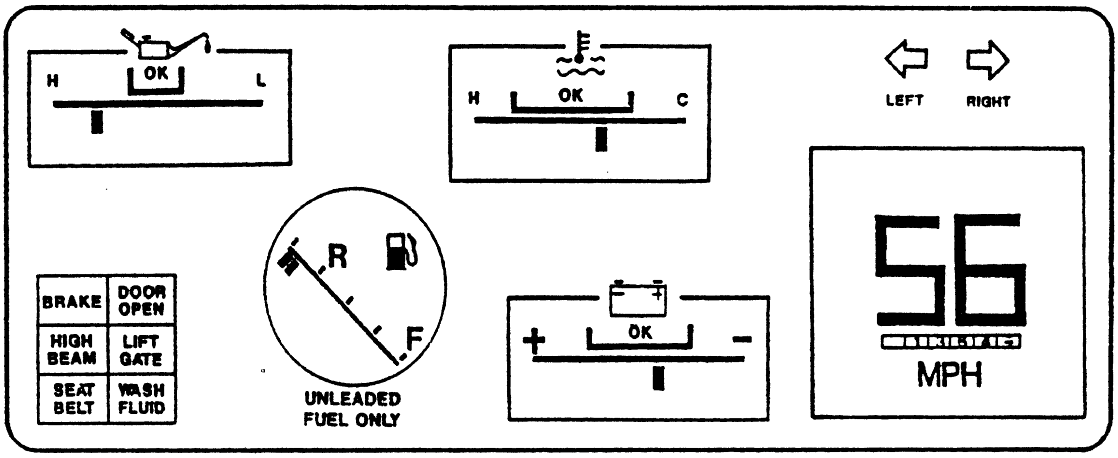
Figure 62. Clusters Examined by Green (1984)



Cluster with zone-coded bar display - Class C3.



Cluster with zone-coded bar displays - Class D1.



Cluster with zone-coded bar displays - Class C2.

(fig 62 continued)

After making decisions using the displays, driver preferences for the various designs were obtained.

With regard to engine gauges, the data showed that they should be moving pointer displays. As an example, the error data for temperature displays is shown in Table 27. Further, the general trend was for displays labelled with "ok" (a range marker) to be best understood. Color coded pointer displays were understood almost as well. On the other hand, how the scale intervals were marked (numbers, words, or abbreviations) did not matter. The critical element was providing scale anchors.

Table 27, Percent Error Data for Engine Temperature Displays

Rank	Labels	Percent Error
1	C-OK-H	19.5
2	140-180-220-260 (color-coded)	21.4
3	C-OK-H	22.3
4	C-H (color-coded)	23.2
5	140-180-220-260 (with OK)	23.8
6	C-H	24.0
7	C-H	28.8
8	140-180-220-260	31.2
9	cold-OK-hot	34.3
10	xxx degrees F (color-coded)	39.6
11	xxx degrees C (color-coded)	49.4
12	xxx degrees F	52.2
13	xxx degrees C	57.1

Source: Green (1984)

Vertical bars indicate p<.05

Many are surprised that drivers understood the temperature displays when numbers were used to label the scale, but not when numbers alone were used. It was clear from the questions of driver knowledge, that many did not have a sense of what normal engine temperature was (or other engine parameters for that matter), but when given a scale, could make a proper decision. (As an aside, color coding numeric displays did make them more understandable than uncoded displays, but they were still not as well understood as simple moving pointer displays.)

All of these differences were found across engine displays. In many cases the differences were statistically significant and, in almost every case, the magnitude of the error differences was large enough to be practically significant.

With regard to the fuel displays, there were few significant differences. In general, moving pointer displays were more likely to be read correctly than numeric displays though there were almost no statistically significant differences within categories. The rank order of those displays based on their error rate is shown in Table 28. The message is that it does not seem to matter how a fuel gauge is labelled (E-F, 0-1, etc.) as long as some anchor is provided that a driver can interpret.

Table 28, Rank Ordering of Fuel Displays

Rank	Label	
1 (best)	0 - 1/2 - 1	gauge
2	0 - F	gauge
3	0 - 1/2 - 1/1	gauge
4	Empty - 1/2 - Full	gauge
5	Empty - Full	gauge
6	R - F	gauge
7	% Full	numeric
8	0 - R - F	gauge
9	0 - 1/1	gauge
10	0 - 1	gauge
11	1/4 - 1/2 - 3/4	gauge
12	0 - 2/4 - 4/4	gauge
13	0 - 4/4	gauge
14	Miles to Empty	numeric
15	E - F	gauge
16	Miles Left	numeric
17	x/8	numeric
18	Liters	numeric
19	x/16	numeric
20	E - 1/2 - F	gauge
21	"moons" 0 - ◒ - ●	gauge
22	Gallons	numeric
23 (worst)	Kilometers to Empty	numeric

** No statistical differences in ranks (very few data points)

Because of its importance, readers are encouraged to retrieve the Green (1984) paper for further details.

Summary

In the general human factors literature, there is one study which empirically examines the benefits of color banding and only one study that examines various types of scale marks other than tick marks. Nonetheless, most of the recommendations for color banding and marking seem to make sense, especially concerning the use of red and green. Other recommendations, however, deserve review. One context in which these recommendations deserve study is across cultures. For example, in China, red is considered positive while in most western countries it is considered negative. The implications of this cultural difference for understanding displays for machinery is unknown. It is, however, becoming increasingly important as products are marketed on a worldwide basis and as language-free methods of coding displays (such as through the use of color) become increasingly common.

Concerning automotive-specific applications, the Green (1984) study is critical. If they are to be understood by the general public, engine displays should be moving pointer displays, the normal zone should be marked (with "ok," "normal", or whatever), that color coding should be considered, but that it matters little how the scale is marked (words, abbreviations, symbols, or numbers). For the fuel display, differences between labelling schemes are small though moving pointer displays tend to be better understood than numeric displays. The method used to collect the data was simple and compelling, and the results are both practically and statistically significant.

HOW SHOULD POINTERS BE DESIGNED?

Several studies have examined how pointers and arrows should be designed. As with many topics covered in this review, most of the early research was quite definitive and there has been little research on this topic since the early 1960's.

White (1951)

This experiment pertains both to pointer alignment cues and to pointer design but was placed here because the conclusions with regard to pointer design are more important. In the first of three experiments, 40 male pilots and cadets served as subjects. The pointer being used at the time extended in both directions from the pivot point, a design which could lead to 180 degree reversal errors. The focus of the experiment was to look at alternative designs of the pointer base to eliminate such errors. The five pointers examined, all 1-1/8 inches long x 3/32 inches wide, are shown in Figure 63.

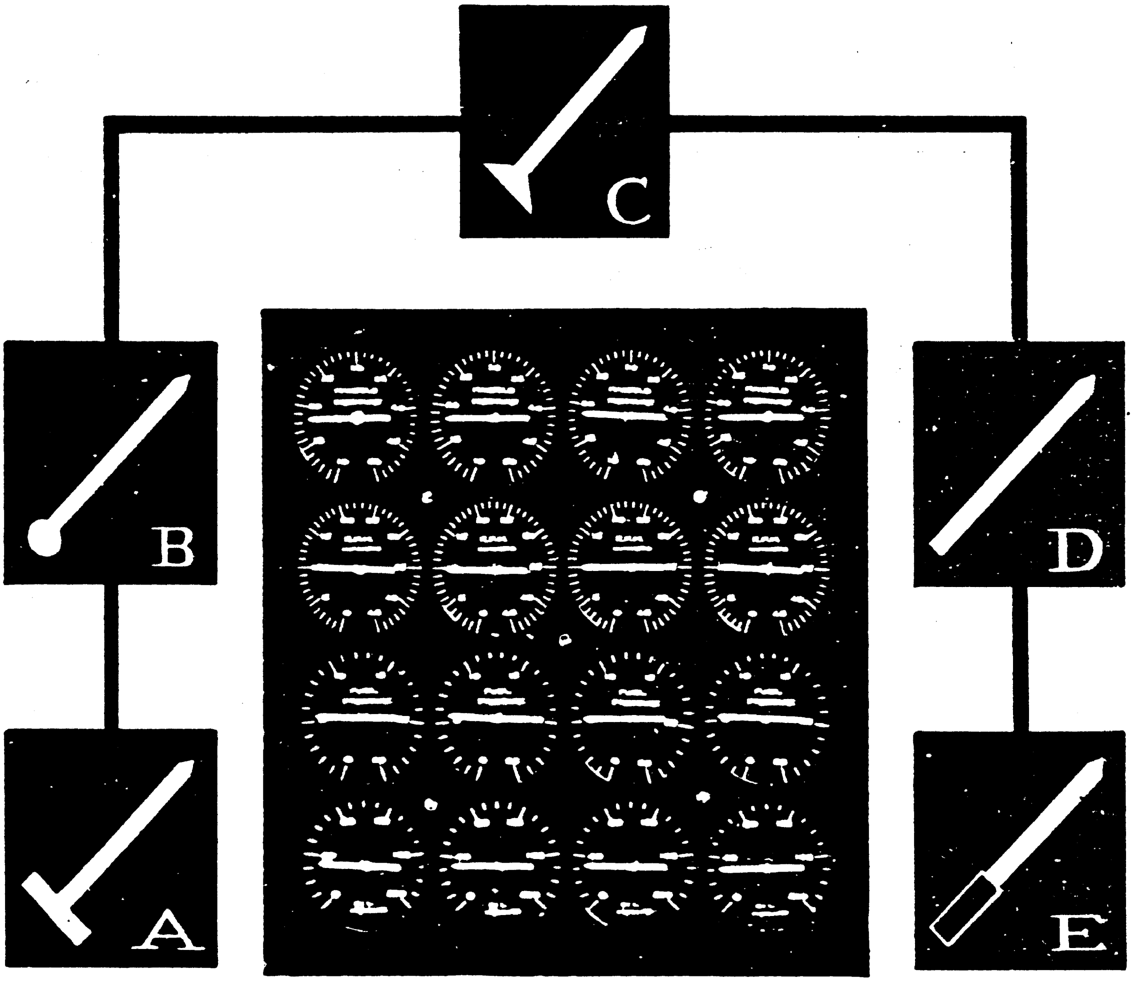


Figure 63. Pointers Examined by White (1951)

Subjects were shown a filmstrip of 4x4 arrays of gauges, viewed at 28 inches. Subjects moved a hand-held toggle switch one way if all gauges were in tolerance, another if any were out. Each subject saw only 1 of the 5 designs.

Table 29 shows the total number of errors, reversal errors, and response times for the 5 pointer types for pilots and cadets. The differences among designs based on either the error or time measures were not significant.

Table 29. Error Data for Pointers from White (1951)

Pointer	-------- Mean # Errors -------				- Mean Time --	
	-- All Types --		-- Reversal --			
	Pilots	Cadets	Pilots	Cadets	Pilots	Cadets
A	3.1	2.2	2.1	1.5	2.00	2.11
B	3.3	2.1	1.9	1.4	1.62	1.91
C	2.8	3.1	1.9	1.1	1.80	2.05
D (current)	2.9	2.5	2.5	1.8	1.73	1.90
E	4.3	---	1.7	---	2.80	----

The second and third experiments, which concern the merits of alignment at various clock positions, are mentioned in the pointer alignment section.

Papaloizos (1961)

This paper describes two very comprehensive experiments concerning a variety of dial design features including pointers. Because it is one of the few studies concerning pointer design, it has been included here rather than in other sections. These experiments were conducted to support the design of a "comparator" shown in Figure 64.

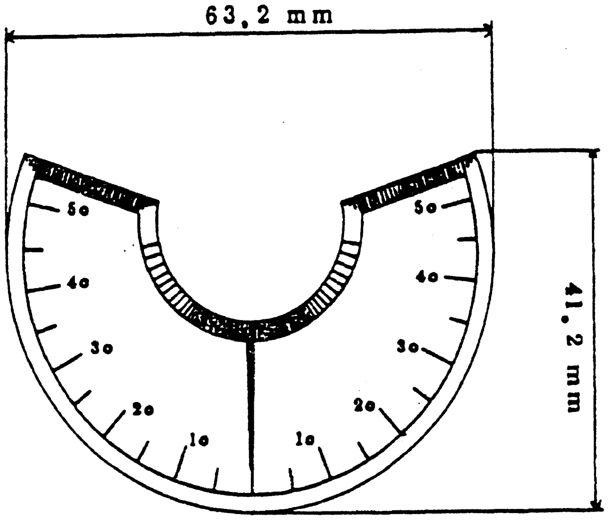


Figure 64. "Comparator" Examined by Papaloizos (1961)

In experiment one the variables manipulated were those associated with the dial frame (saturation of the grey-light or dark, upper arc of the window-small or large, finish-matt or glossy), width of the pointer head-thick (same width as major graduations) or thin (same width as minor graduations), length-long (overlapping minor graduations) or short (touching minor graduations), color-same or different from major graduations, graduations (colors of majors and minors same or different, colors of minors one or two colors, minors all same height or different heights (staircase), and dial color (light or dark saturation, hue=neutral, yellow, green, or blue). Munsell designations for the various colors and other details concerning the various marking sizes are given in the paper.

Eight people were shown dials at 30 cm for 200 ms each in a randomized fashion. Their task was to read the dial. A total of 12,800 readings were obtained, of which about 20% were errors. In general, none of the factors associated with the dial frame were significant though the interation between saturation and finish was. The light glossy grey on a dark matt background was best. For pointers, none of the main effects were significant, though again an interaction was found. (The long hand-same color and short hand-different color combinations were better than long hand-different color and short hand-same color.) No explanation for this interaction is provided and it doesn't seem to make sense, unless somehow the results obtained were influenced by the color of the minor marks. On the other hand, the color of the graduations did matter; making the minor graduations a different color led to significantly fewer errors. The dial color and saturation affected performance. Readers should see the original paper for the particular combinations that did well and poorly.

In the second experiment 16 dial designs were examined, formed from combinations of 8 types of graduation marks and two pointer designs. (See Figure 65.) Sixteen people read each dial 20 times. Dials were projected on a screen until the participant pressed a button to close the shutter.

There were no significant differences in the reading times for the two pointer types. There were, however, significant differences between display designs. In Table 30, differences of 60 ms were associated with significant differences at the p=.05 level. Notice that reducing the number of tick marks (designs G and Z) made the dials more difficult to read.

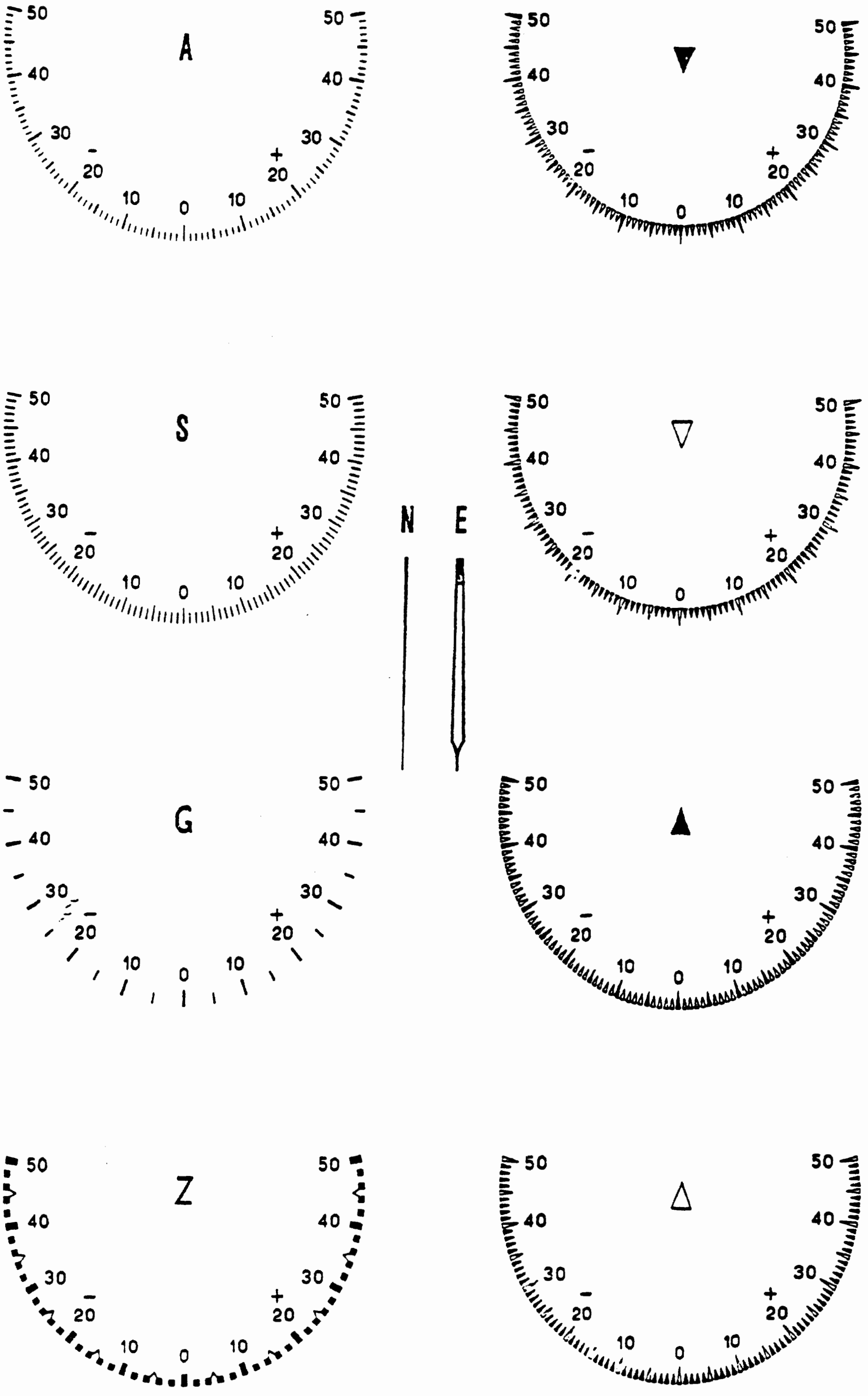


Figure 65. Graduation Marks Examined by Papaloizos (1961)

Table 30, Response Times to Dials in Papaloizos (1961)

Design (Figure 64)	Mean Response Time (ms)
A	1647
V (dark)	1648
^ (dark)	1685
V	1785
^	1795
S	1832
G	1886
Z	2421

vertical bars indicate p<.05

While a detailed discussion of errors is not provided for the second experiment, there was a high correlations between the two measures (r=.79) suggesting those data support the conclusions drawn from the response time data.

Spencer (1963)

This paper reviews the literature on the design of pointers for instrument scales. Since most of the studies he reviews are also reviewed here, thoughts about his review will be confined mainly to his recommendations.

Spencer begins his review by defining some terminology. He notes that for every pointer design the visible length, the visible tail size, the scale radius, and the boss diameter (around the pivot point) must all be given. Spencer notes that there is controversy as to which parts of the pointer are most important. The suggestion is that there is no difference between outer half length and full length pointers. It is not clear if full length refers to pointers with a balancing tail, or ones that extend fully to the scale baseline, though he does comment that pointers up to the radius of the scale are better than shorter pointers. Width and shape of the pointer have little effect on performance provided the tip is pointed and the pointer contrasts well with the instrument face. He concludes pointer color seems to have little effect on performance. The review of Spencer is very difficult to summarize and readers interested in that topic should retrieve his article.

Arrows for Other Than Dials

The most common use of pointers is as indicators on signs (Bartlett and Mackworth, 1950; De La Mare and Walker, 1962; Bryant and Smith, 1976; Gordon, 1981). The results from those studies are of limited use here as the major design factor is the type of head provided. In signing studies, people are shown a series of arrows, often ones that are small, and asked to indicate the direction in which they are pointing. Hence, the context is one of deciphering the direction when the arrow is small. Such emphasis encourages the use of large heads. For dials, such designs are poor because they cover the numbers to be read.

Summary

1) The variations in pointer design (length, head shape, etc.) one normally finds among gauges have only a small effect on performance.

2) It is clear that the pointer should be narrow near the tick marks and close in width to the tick marks.

3) The literature also suggests that the pointer should be a color that contrasts well with the scale face and the scale marks, but that the benefits of making it a different color depend on whether the scale marks are all one color. If the scale marks are of multiple colors, then making the pointer a different color will be of minimum benefit.

4) It is not clear how long a pointer should be. Based on the work of Vernon, 1946 (described in another section), it appears the pointer should almost touch the scale marks when quantitative reading is required. Gaps of even .25 inches should have a very minor effect on performance. However, for check reading, the Oatman work suggests that pointers should be as long as possible.

5) There is little evidence concerning tail length and it probably doesn't matter very much. Tails are desired for mechanical balance, and based on the Oatman work, should assist in check reading. On the other hand, they cover part of the scale face and large tails can lead to reversal errors.

WHAT REALLY MATTERS IN READING GAUGES?

This section discusses three very different studies that try to assess the relative importance of various design considerations. Again, designers are always faced with making tradeoffs and, they need to know where to make design compromises that will have minimal impacts on the legibility or understandability of displays.

Control Systems Laboratory (1955)

This obscure and theoretical study examines, in a methodical way, several factors affecting the reading of dials. The study takes an information theory perspective. In brief, that perspective views the number of choices available at any given moment as measurable in bits, where the precise value depends upon the probability of each alternative. Those seeking further information on this perspective should consult a psychology text written from that perspective (e.g., Garner, 1975) or an electrical engineering text concerning communications theory.

In the first experiment nine people were presented a variety of dial-like images shown for 100 ms, with each person usually seeing just one design. (See Figure 65.) The person identified the sector indicated by the pointer.

The measures used in this experiment were input information (H-in), the uncertainty of the pointer position measured in bits, and information transmitted (T-in,out), a measure of the overlap between input and output (also measured in bits). In Table 31 notice that error-free performance is about 12 sectors for unmarked dials and almost 16 for marked dials with peak performance being about 3.8 bits for unmarked dials and 4.3 bits for marked ones. Thus, for simple check-reading tasks, it is wise keep to the number of critical categories down to about 16 for circular dials. For semicircular and arc meters, it seems likely the number of categories should be reduced in proportion to the relative angular size.

Table 31, Dial Reading Responses-# Sectors for Single Dial

# divisions	H-in (bits)	T-in,out (bits)	
		----- dial ------	
		blank	black and white
12	3.58	3.6	3.6
16	4.00	3.8	3.9
24	4.58	3.7	4.3
32	5.00	3.7	4.1
36	5.16	3.8	4.3
48	5.58	---	4.1

Source: Control Systems Laboratory (1955)

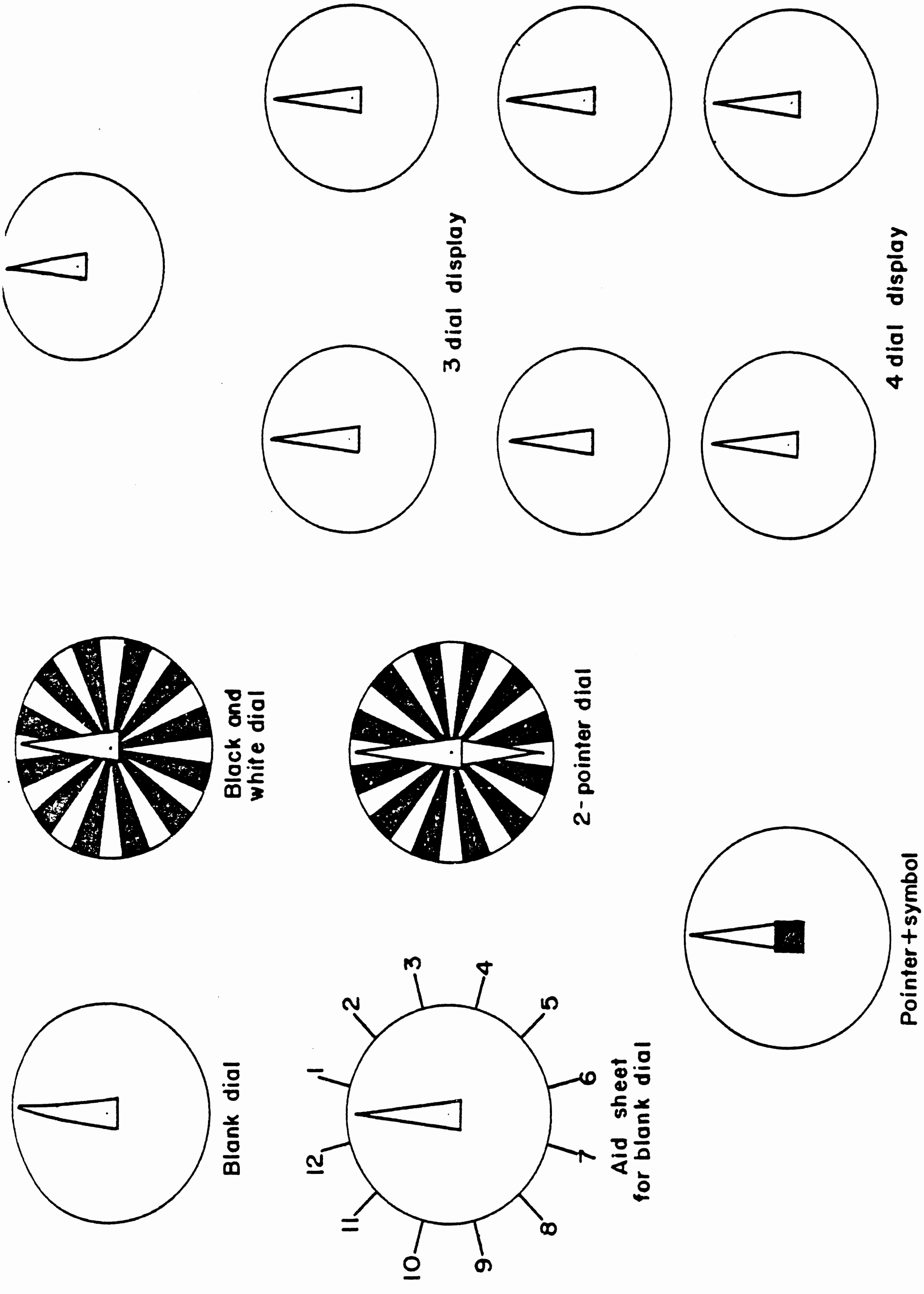


Figure 66. Dials examined by Control Systems Laboratory (1955).

Shown in Table 32 are the data for multiple dial conditions. As the number of dials increases, so too does the total information transmitted, but the information per dial drops off. These numbers suggest that the amount of information one can gain from an array of dials in a single glance is limited.

Table 32, Information Transmitted vs. # of Dials

# Dials	T-in,out/dial (bits)	T-in,out Total (bits)
1	4.3*	4.3
2	3.2**	6.3
3	1.9**	5.6
4	2.1**	8.6

Source: Control Systems Laboratory (1955)

* from Table 31
** numbers are rounded off

Also examined were ways to increase the information transmitted per dial (placing multiple pointers of contrasting colors on single dials, adding letters and numbers in the center). In the alphanumeric condition, participants identified the pointer position and recalled the character. As shown in Table 33, adding such information raised the information transmitted considerably, and there is no evidence of saturation due to added symbols. Interestingly, performance with 2 dials and a symbol (6.8 bits) was not as good as performance with 2 dials alone (6.9 bits).

Table 33, Information Transmitted from Single Dials with Multiple Pointers of Multiple Dials with Symbols

# pointers	H-in (Input Information)
2	7.7
3	8.9 bits/display

Display	Total H-in (bits)	T-in,out (bits)		Total T-in,out (bits)
		pointer	symbol	
1 dial + 1 symbol	9.6	4.3	4.3	8.6
1 dial + 2 symbols	14.6	4.3	9.5	13.7
1 dial + 3 symbols	19.6	4.0	13.6	17.6
2 dials + 1 symbol/dial	19.2	1.5	5.3	6.8

Source: Control Systems Laboratory (1955)

A similar set of experiments was conducted using straight scales. In that set of experiments both blank scales and strips of 3 to 36 squares (in a checkerboard pattern) were tested. The marker (pointer) was a either a red chip, or a black chip with a white cross when scales were color-coded. As before, participants indicated where the pointer had appeared. Table 34 contains the results.

Table 34, Information Transmitted from Single Vertical or Horizontal Scales

# Divisions	H-in (bits)	T-in,out (bits) Scale Blank Vert	Blank Hor	Black & White Vert	Black & White Hor
12	3.58	3.1	2.8	3.2	3.6
16	4.00	3.0	3.3	3.6	4.0
20	4.31	3.1	3.3	3.2	4.0
24	4.58	2.5	3.4	2.8	3.7
28	4.80	2.7	3.2	2.9	3.7
32	5.00	3.1	3.3	2.5	3.7
36	5.16	2.6	3.1	2.6	3.6
mean		2.9	3.2	3.0	3.8

Source: Control Systems Laboratory (1955)

The maximum amount of information transmitted was typically 3.4 bits for blank scales and 4.0 bits for marked scales. There was a slight tendency (probably not significant) for horizontal scales to transmit more information. These figures are slightly less than those for circular dials (3.6 and 4.3 bits respectively).

As with the dials, the benefits of adding color coding to strips of 16-40 squares (using the same color to codes groups of 6-8 squares) was examined. Adding color coding raised the information transmitted by about 1/2 bit. (See Table 35.)

Table 35, Color Coding of Horizontal Scales

Scale	Average T-in,out (bits)
black & white	3.8
1 color/6 squares solid & hatched	3.3
1 color/8 squares solid & hatched	3.7
1 color/6 squares solid only	4.2

Source: Control Systems Laboratory (1955)

Also examined was the effect of adding letters or numbers to the scale as was done with dials. For 24 square strips adding a character to the center of the display added only slightly to the information transmitted. Performance dropped off as the character was moved to the end of the scale reaching approximately 0 when 5 or more squares from the middle of the display. For dials, part of the pointer is in the center, so there is an advantage of looking there for the pointer and the character displayed. For linear scales, the two pieces of information are in different locations and it is difficult to absorb them within the brief exposure.

Paralleling the dial work, multiple linear scales were also examined, though only for color-coded scales. Scales with 24 squares (H-in=4.58 bits/scale) were examined, using 4 colors to partition the scale into groups of 6 squares (the optimum for a single strip). For 2 scales the amount of information transmitted was 2.5 - 2.7 bits or 5.0 - 5.4 for each pair. For 3 scales, the amount of information transmitted was 1.6 - 1.9 bits/scale or 4.8 - 5.7 bit for each triple, no more that the amount for 2 scales.

This low amount of information transmitted/scale in the 3-scale condition led the authors to explore conditions where there were up to and usually fewer than 24 graduations/scale (3, 4, 6, 8, 12, 18, and 24). Shown in Table 36 are the results for cases where there are up to 8 squares/scale. Data have been pooled across the number of scales to simplify the table. Information transmitted peaked at about 12 bits. In each case for combinations of scales and number of squares/scale at the maximum, adding more scales decreased information transmitted.

Table 36, Information Transmitted for Multiple Scales, 3-8 Squares/Scale

Squares /Scale	# Scales	H-in (bits)	T-in,out (bits) Scales			
			1-4	5-7	8+	All
3	4	6.3	6.2			6.2
	5	7.9	5.7	1.5		6.2
	6	9.5	5.7	2.5		8.2
	7	11.1	5.6	4.2		9.3
	8	12.7	5.4	3.4	1.0	9.8
	10	15.8	3.8	1.6	0.8	6.2
4	4	8.0	7.8	7.0		7.0
	5	10.0	7.3	1.6		8.9
	6	12.0	6.6	2.6		9.2
	7	14.0	6.9	4.2		11.1
	9	18.0	6.4	2.8	0.7	9.9
	10	20.0	6.9	2.8	1.7	11.4
	12	24.0	5.3	1.6	2.5	9.4

Table 36 (continued)

Squares /Scale	# Scales	H-in (bits)	T-in,out (bits) Scales			
			1-4	5-7	8+	All
6	4	10.3	9.1			9.1
	5	12.9	8.5	0.9		9.4
	6	15.5	8.5	3.2		11.7
	8	20.7	7.3	3.4	0.5	11.2
	10	25.8	8.6	2.8	2.5	13.9
8	4	12.0	10.7			10.7
	6	18.0	9.7	3.7		13.4
	8	24.0	8.0	2.3	0.7	11.0
	10	30.0	10.3	1.6		11.9

They summarize their data from the various combinations of conditions in Figure 67.

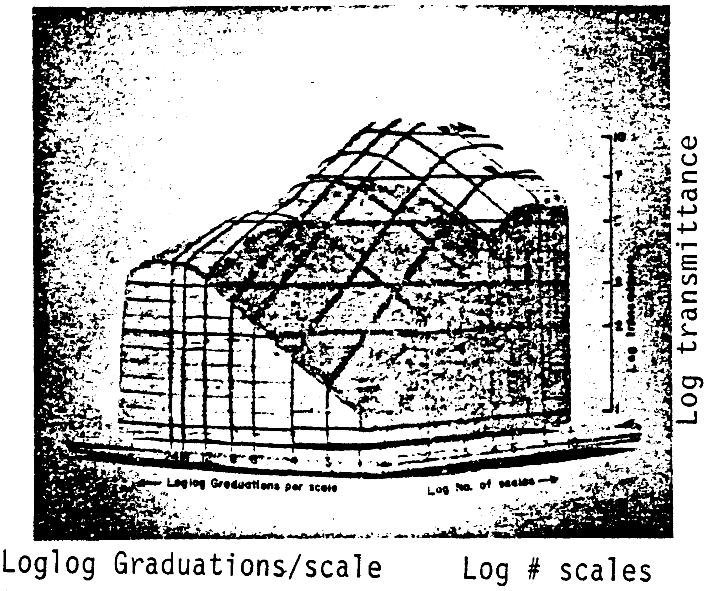


Figure 67. Smoothed Data from Control Systems Laboratory (1955)

Groth and Lyman (1961)

This abstract experiment examined the usefulness of various features of dials. Features included two pointer tips (clock hands) in an unstructured field, the addition of x and y axes, enclosure of pointer tips by a perimeter, heavy dots to emphasize the 3, 6, 9, and 12 o'clock positions, blank background, cross-hatched background, and dotted background.

Twenty-one students were given 24-page booklets. On each page were 12 dial-like displays. (See Figure 68.) Their task was to record the time shown. They had 48 seconds to complete each page. The paper and pencil method chosen is an interesting contrast to the methods used in other studies in this section.

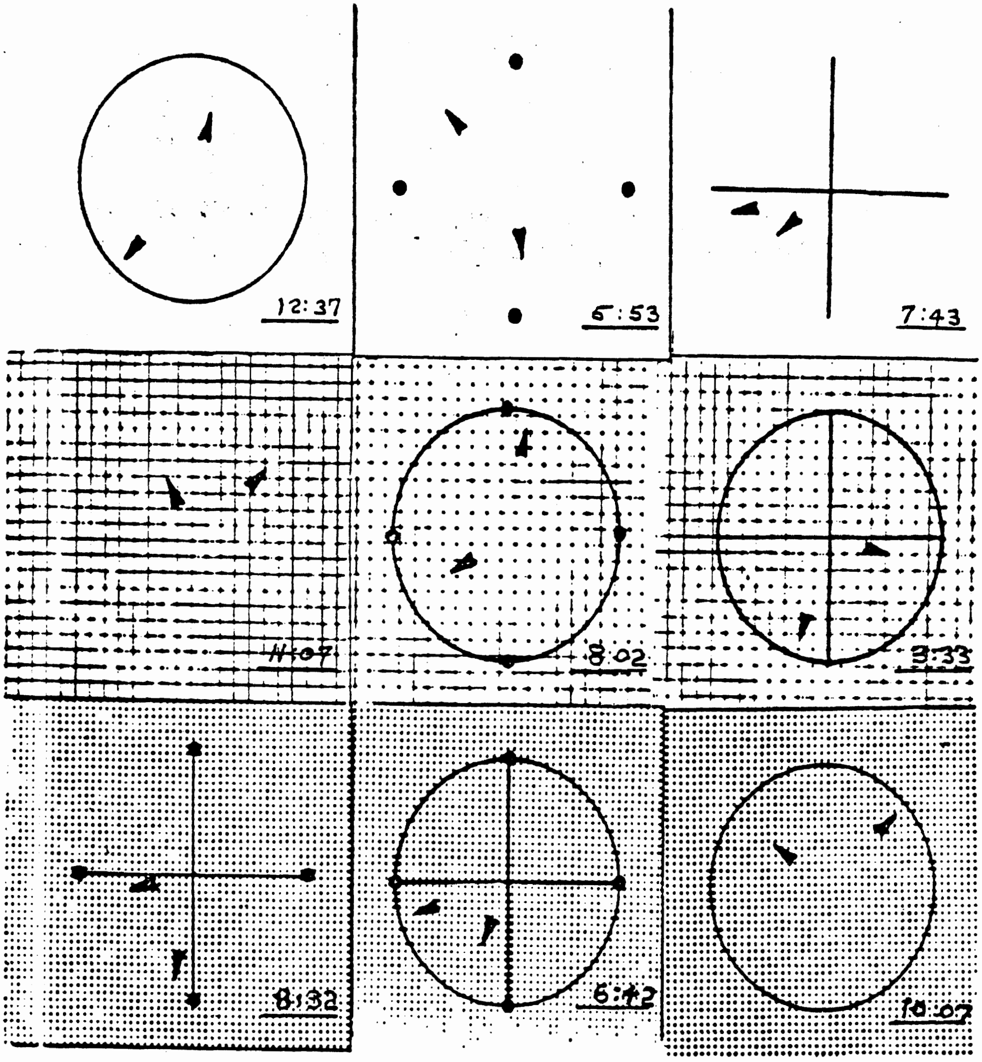


Figure 68. Displays Examined by Groth and Lyman (1961)

Shown in Table 37 are the performance data averaged across backgrounds. (Surprisingly there was no differences in backgrounds.) One clear point that emerges from this table is that the perimeter is a critical element of a dial. In many contexts dials are stylized and scale baselines are not provided. These data suggest that such designs are less likely to be read accurately than designs in which baselines are provided.

Table 37. Rank Ordering of Various Dial Features (Groth and Lyman, 1961)

Feature Group	Correct Responses		Completed Responses	
	Rank	%	Rank	%
Pointers+Perimeter + Dots	1	56.6	1	85.3
Pointers+Axes	2.5	55.9	2	84.9
Pointers+Perimeter + Axes	2.5	55.9	3	83.6
Pointers+Dots+ +Perimeter+Axes	4	52.5	6	80.6
Pointers+Perimeter	5	51.7	4	82.2
Pointers+Dots + Axes	6.5	48.3	7	80.3
Pointers+Dots	6.5	48.3	8	77.9
Pointers	8	40.0	5	81.3

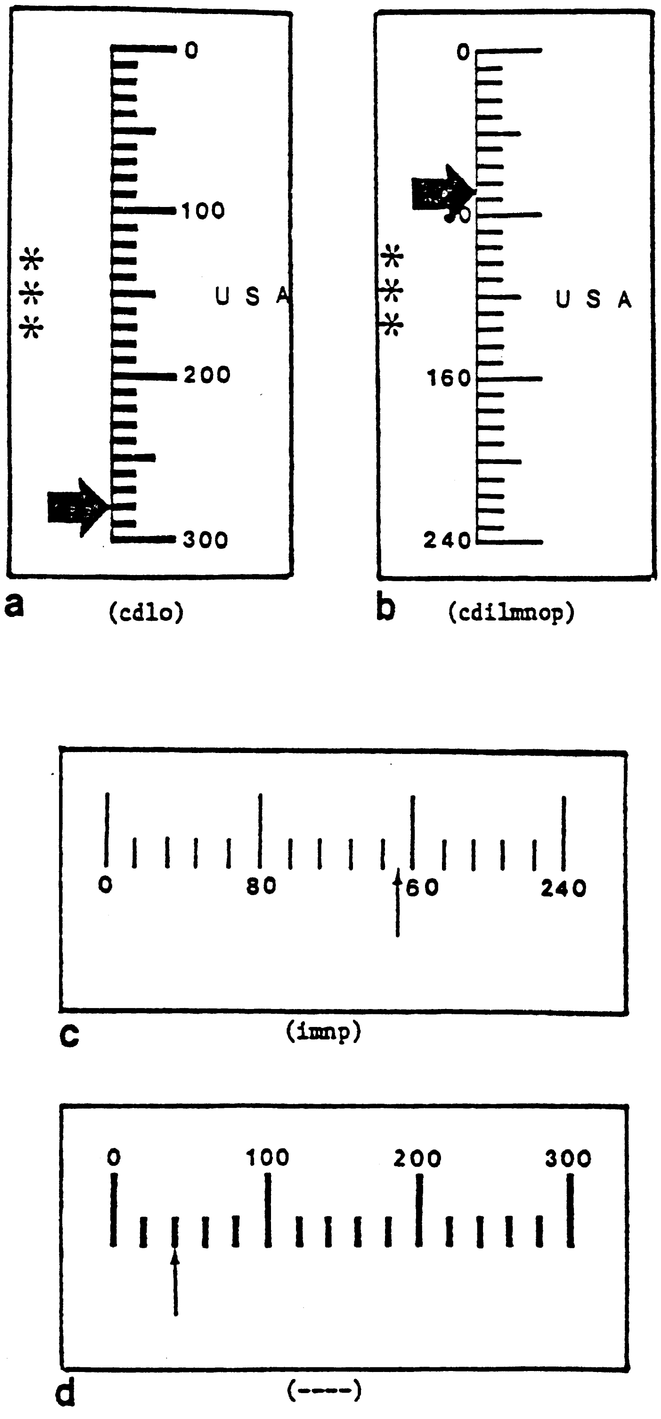
Whitehurst (1982)

This is probably the only study that takes a practical approach to examining the relative importance of various gauge design considerations. Two groups of 16 students were tested using a complex experimental design. One group saw black on white scales, the other white on black. The other factors varied and levels selected are shown in Table 38.

Table 38. Conditions Examined by Whitehurst (1982)

Factor	Level "Good"	Level "Bad"
Clutter	none	3 asterisks and a USA label
Pointer	skinny (.8mm)	fat (6.4mm) width shaft
Interpolation	no	yes-required between marks
Marker Width	.8 mm	1.6 mm (double recommended level in HF textbook)
Scale Number Location	opposite ptr	same side of scale as pointer
Scale Orientation	vertical	horizontal
Numerical Progression	10's or 20's	8's or 16's

On each trial participants pushed a button once to open the slide projector shutter and start a timer, and a second time to stop the timer when they were able to read the test display. The reading was then recorded. Each of the 32 combinations of the various factors appeared twice in the sequence with each of the two appearances of each combination showing different values. Samples of the displays tested appear in Figure 69.



a (cdlo)

b (cdilmnop)

c (imnp)

d (----)

Figure 69. Some Displays Tested by Whitehurst (1982)

In this experiment there were no statistically significant differences in the time to read black-on-white vs. white-on-black scales (8.10 vs 8.05 s), but significantly more errors were made in reading white-on-black dials (14.8 vs 9.9%). Dials were read significantly faster when scales were numbered in 10's or 20's as opposed to 8's or 16's (5.4 vs 9.8 s), when interpolation was not required (6.9 vs 9.1 s), and when the scale markers were widely spaced (7.2 vs 9.0 s).

Shown in Table 39 is the variance accounted for by each factor in the ANOVA of the data. Except for numerical progression and interpolation, most of the factors accounted for only a very small part of the variance. This could imply those factors have only a minor impact on real user performance and are unimportant. While these data do suggest a possible rank ordering of design factors, this author is not convinced of the previous conclusion for the following reasons.

Table 39. Percent Variance Accounted For

Source	RT Mean	Error Mean
Numerical Progression	41	15
Interpolation	9	9
Scale Unit Length	4	4
Scale Orientation	1	1
Marker Width	1	1
Clutter	0	1
Pointer Design	0	0
Scale Number Location	0	0
2-way Interactions	4	10
Subjects	18	7
Residual	15	10

Source: Whitehurst (1982)

First, the response times reported seem extremely long. This author has conducted a large number of studies involving reading speedometers and in them, times on the order of a second or so are common. While admittedly the participants had very little practice here (but a fair amount of practice in the speedometer studies), the differences still cannot be reconciled.

Second, the levels selected for the various factors don't seem to represent equal degrees of "good" and "bad" (relative to the range available). For example, the good and bad values in marker width differed by only 2:1, but the marker width differed by 8:1. In fact, if the figure in the paper is indeed indicative of what the test display looked like, the "good" pointer looks too thin. With regard to scale orientation, the difference (which is actually to relative location of the scale marks and the pointer) seems confounded with the scale baseline orientation. Finally, it seems apparent that larger differences due to progression would have occurred had a more difficult progression scheme been used (e.g., 7's instead of 8's).

Summary

1) What matters most? While not discussed in this section, it is abundantly clear that the match of the display with the task is of critical importance. Numeric displays are preferred for quantitative reading, moving pointer displays for check reading, and so forth. For check reading tasks, aligning pointers can lead to large reductions in reading time and errors.

With regard to the design of pointer displays, the conclusions about what matters most are based primarily on the Whitehurst (1980) study. While there are concerns about how the study was conducted, his results fit in with conclusions reached by others. In brief, the key factors seem to be numerical progression and, to a lesser extent, the degree to which interpolation is required. Other studies (e.g., the Control System Laboratory 1955 work) suggests that color coding is also a key factor. In that study adding color coding to a dial increased information transmitted almost as much adding marks to it. Other studies (Groth and Lyman, 1961) suggest that the presence of a baseline is important as well.

2) What factors are of secondary importance? Based on the Whitehurst work and other studies, those factors include marker length and width, marker separation, pointer length and width, scale orientation, clutter and background design, and scale number location.

DESIGN RECOMMENDATIONS

This section contains a short list of recommendations for the design of automotive displays, in particular gauges, based on this literature review. The goal of these recommendations is to help designers develop displays that are easy to use. Here, easy to use displays are both legible and understandable. Readers seeking a more extensive list of recommendations should consult the Mil Handbook (U.S. Department of Defense, 1981a) or the Mil Standard (U.S. Department of Defense, 1981b).

Display Type

Rule 1. The particular type of display that is easiest to use depends on the task for which it is intended. In general when an exact number is required, a numeric display should be provided. When the primary task is check reading, a moving pointer display is best.

Recommendation 1: For motor vehicles a numeric display is preferred for the speedometer.

Recommendation 2: The engine and fuel gauges should be moving pointer displays.

Display Format

Rule 2. When a group of moving pointer displays are to be check read, they should be arranged so their pointers are aligned when they all show normal values.

Comment: There is considerable discussion in the literature as to which position is best. The key is consistency, with alignment at 9 or 12 o'clock being most common for circular displays. For arc meters and horizontal and vertical scales, alignment of pointers is straightforward.

Recommendation 3: When more than one engine gauge is on the instrument panel, they should be close to each other and arranged so their pointers are aligned when all show normal values. They should not be grouped with the fuel gauge.

Rule 3. For quantitative reading, the ranking of moving pointer displays from best to worst is: circular, arc, horizontal, vertical.

Comment: This rule is based on laboratory data in which visual search is not required. It assumes that reading time is strongly influenced by how far the pointer tip is from the fixation point.

Rule 4. For check reading, the differences between moving pointer displays are small with vertical displays tending to be more difficult to read.

> Comment: Both quantitative and check reading performance are markedly affected by the compatibility of the pointer motion and the associated response. So, for example, a design that required users to move a switch up when the pointer moves down would be a poor.

Scale Marks

Rule 5. Scale marking considerations are less important than the choice of the proper display and the alignment of pointers for check reading.

Rule 6. For zero-based numbers, scales should be marked with values greater than 1 and numbered in even multiples of 10 (0, 10,... or 0, 100,..., etc.) when an exact number is desired. Easy to use displays are ones which minimize the number of mental operations a viewer is required to complete to interpret them. Nondecimal schemes (0, 2.5,..., or 0, 1.7, 3.4,..., etc.) are much more difficult for people to understand.

> Recommendation 4: Moving pointer speedometers for production vehicles should be numbered in increments of 10, not 20 mph. It is not clear, however, if numbers should be associated with fives or tens. Numbering the tens is compatible with how people process numbers but incompatible with the way speed limits are posted (35 mph, 55 mph, etc.). Research to address this question should be conducted.
>
> Comment: Recommendation 1 takes precedence over Recommendation 4. Numeric speedometers are preferred over moving pointer speedometers.
>
> Comment: Many automotive engine displays are not zero-based for normal operation. For example, engine temperature displays (which are usually check read) almost never show values between 0 and 140 degrees when the engine is running. The same is true for electrical system voltage, which is invariably 13.5 volts, plus or minus 2.5 volts. Therefore, this rule does not apply to these displays.
>
> Recommendation 5: If labelled with numbers, other engine gauges (oil pressure, oil level, electrical system current) should have the zero point labelled and numbered with 1's or 10's as appropriate.

Rule 7. Scale marks should be provided down to the level to which a display must be read. If a speedometer is to be read to the nearest mph, then marks showing the units should be provided. If it is read to the nearest 5 mph, then only marks at that level should be provided. It is not clear how accurately speedometers are read. That issue should be investigated experimentally.

Rule 8. Scale marks should not appear at noninteger points on a scale unless the values being displayed are not integers.

> Comment: According to this rule, tick marks on a speedometer in 2.5 mph increments (i.e., halfway between 5 and 10) are ill advised. Noninteger markings add to the mental effort required to read a display and make it more difficult to read.

Rule 9. For displays that are check read, how a scale is marked once the normal range is shown tends not to be important.

> Recommendation 6: Provided normal is clearly shown (e.g., range marks labelled with "ok"), it does not matter how engine gauges are labelled. Numbers, words, abbreviations, and symbols are equally informative, and any of them can be used.

Rule 10. For qualitative readings two anchors may be sufficient if the measured dimension is well understood by viewers.

> Recommendation 7: It does not matter much how a fuel gauge is labelled. Drivers understand most of the common labels (E - F) and even some of the uncommon (0/4 - 4/4) labels nearly as well.

Rule 11. Dials should have breaks between 0 and the maximum. It is not clear how big they should be.

Rule 12. Marked intervals should be at least 1/2 inch apart. (Marked intervals are those points on the scale that have numbers shown next to them, e.g., 0, 10, 20, etc.). Some have argued for intervals of an inch or more.

Rule 13. Scale marks should be separated by at least 1/10 inch.

Rule 14. Scale marks (and pointers) should be at least .03 inches wide. Should wider marks be used, they should always be considerably less than the gap between marks to avoid figure ground reversal problems.

Rule 15. It doesn't matter if scale numbers are on the inside or outside of a dial or on the same or different side of the scale as the pointer.

Rule 16. Nonstandard marking schemes (staircase tick marks, using log scales to linearize data) offer minor, if any, performance benefits.

Zone Markings

Rule 17. Zone markings ("ok," "normal," etc.) should be provided on displays which are check read. They make displays much easier to check read. Labelling them with words or color bands is about equally effective.

> Recommendation 7: Every engine gauge should have zone markings.

Rule 18. Normal zones should be colored green. Danger zones should be colored red. There is debate as to whether other zones should be white or yellow.

Pointers

Rule 19. For electronic displays where multiple segments are used to represent a pointer, only a single segment should be illuminated. (A cursor design is easier to understand than a fill design.)

> Comment: This rule has been experimentally verified for engine displays but not for speedometers. Of the rules listed for pointer design, this one is likely to have a major influence on performance and is an exception to Rule 4.

Rule 20. The gap between pointer tips and the associated tick marks should be between 0 and 1/4 inch for accurate quantitative or qualitative reading. For check reading of multiple aligned displays, longer pointers should be provided.

> Recommendation 8: For speedometers and fuel gauges, the gap should be 1/4 inch or less.
>
> Recommendation 9: When multiple gauges for engine functions (temperature, oil pressure, etc.) are provided, longer pointers should be provided.

Rule 21. The pointer width near the tip should be about equal to minor mark width.

Rule 22. If scale marks are all one color, the pointer should be a different color that contrasts well with the marks and the background.

> Comment: If the tick marks are multiple colors, this rule may not hold. This should be investigated experimentally.

The author hopes that this set of rules proves useful to designers. Those with suggestions for additional rules or those seeking clarification are encouraged to contact the author. This set should be viewed as a initial attempt at developing rules for automotive applications, not a final definitive set.

- DESIGN RECOMMENDATIONS -

REFERENCES

Anonymous (1981). Instrument Panel Analyzed, Automotive Engineering, October, 89(10), 79-81.

Armour, J.S. (1984). Speedometer Reading Times (working paper 59). Crowthorne, U.K.: Transport and Road Research Laboratory, as cited in Anonymous (1985), Speedometer Reading Times, Automotive Engineer, February/March, 10(1), 57-58.

Bailey, R.W. (1982). Human Performance Engineering: A Guide for System Designers. Englewood Cliffs, N.J.: Prentice-Hall.

Baines, P.A., Spicer, J., Galer, M.D. and Simmonds, G. (1981). Ergonomics in Automotive Electronics (paper C166/81), Proceedings of the Third International Conference on Automotive Electronics, Warrendale, PA: Society of Automotive Engineers.

Baker, C.A. and Grether, W.F. (1954). Visual Presentation of Information (WADC technical report 54-160). Wright-Patterson Air Force Base, OH: U.S. Air Force Wright Air Development Center, August.

Bartlett, F. and Mackworth, N.H. (1950). Planned Seeing (Air Ministry Publication 3139a). London, England: Her Majesty's Stationary Office.

Bauer, R.W., Cassatt, R.K., Corona, B.M., and Warhurst, F.Jr. (1966). Panel Layout for Rectilinear Instruments. Human Factors, October, 8(6), 493-497. (also released as U.S. Army technical memorandum 4-68, Aberdeen Proving Ground, MD: U.S. Army Human Engineering Laboratories, January).

Boff, K.R. and Lincoln, J.E. (eds.) (1988). Engineering Data Compendium: Human Perception and Performance (volumes 1-4). Wright-Patterson Air Force Base, OH: U.S. Air Force Harry G. Armstrong Aerospace Medical Research Laboratory.

Boff, K.R., Kaufman, L., Thomas, J.P. (1986). Handbook of Perception and Human Performance (volumes I and II). New York: Wiley.

Bryant, J.F.M. and Smith, G. (1976). The Shape of Traffic Signal Arrows, Australian Road Research Board Proceedings, 8, 36-45.

Carr, W.J. and Garner, W.R. (1952). The Maximum Precision of Reading Fine Scales, Journal of Psychology, 34, 85-94.

Chapanis, A. (1960). On Some Relations Between Human Engineering, Operations Research and Systems Engineering (report 8). Baltimore, MD: The Johns Hopkins University, Psychology Department.

- *REFERENCES* -

Chapanis, A. (1965). Man-Machine Engineering. Monterey, CA: Brooks/Cole.

Chapanis, A. (1987). Visual Displays I and II. Lectures 4 and 6 in Pew, R.W. and Green, P. (1987), Human Factors Engineering Short Course Notes. Ann Arbor, MI: The University of Michigan Chrysler Center for Continuing Engineering Education.

Chapanis, A. and Scarpa, L.C. (1967). Readability of Dials at Different Distances with Constant Visual Angle, Human Factors, October, 9(5), 419-426.

Chapanis, A., Garner, W.R., and Morgan, C.T. (1949). Applied Experimental Psychology. New York: Wiley.

Christensen, J.M. (1948). The Effect of Staircase Scale on Dial Reading Accuracy (memorandum report MCREXD-694-1-P). Wright-Patterson Air Force Base, Dayton, OH: U.S. Air Force Aero Medical Laboratory.

Churchill, A.V. (1956). The Effect of Scale Interval Length and Pointer Clearance on Speed and Accuracy of Interpolation. Journal of Applied Psychology, 40(6), 358-361.

Churchill, A.V. (1959). Optimal Interval Length for Visual Interpolation: The Effect of Viewing Distance. Journal of Applied Psychology, 43(2), 125-128.

Churchill, A.V. (1960). The Effect of Pointer Width and Mark Width on the Accuracy of Visual Interpolation. Journal of Applied Psychology, 44(5), 315-318.

Cohen, J. and Dinnerstein, A.J. (1958). A Comparison of a Linear Scale and Three Logarithmic Scales on the Time for Check Reading (WADC technical report 57-63). Wright-Patterson Air Force Base, OH: U.S. Air Force Wright Air Development Center, March, (available from NTIS as AD 118017).

Cohen, V. (1971a). A Psychologist's Laboratory Approach to a Human Factors Problem. in Bernotat, R.K. and Gartner, K.P. (1971). Controls and Displays (The Proceedings for an Advanced Study Institute). Amsterdam: Swets and Zeitlinger.

Cohen, V.V.R. (1971b). Short-Term Memory for Quantitative Information from Three Kinds of Visual Displays (technical report 28). Ann Arbor, MI: The University of Michigan Human Performance Center.

Connell, S.C. (1950). Some Variables Affecting Instrument Check Reading (technical report 6024). Wright-Patterson Air Force Base, Dayton, OH: U.S. Air Force Air Materiel Command, Aero Medical Laboratory.

Control Systems Laboratory (1955). Human Performance in Information Transmission, Part III (report R-68). Urbana, IL: University of Illinois, Control Systems Laboratory, Bio-Systems Group.

Dashevsky, S.D. (1964a). Check-Reading Accuracy As a Function of Pointer Alignment, Patterning, and Viewing Angle. Journal of Applied Psychology, 48(6), 344-347.

Dashevsky, S.D. (1964b). Combining Check-Reading Accuracy and Quantitative Information in a Space-Saving Display. Journal of Applied Psychology, 48(6), 348-350.

De La Mare, G. and Walker, J. (1962). The Visibility of Direction Indicators. Ergonomics, 5(4), 573-579.

Elkin, E.H. (1959). Effects of Scale Shape, Exposure Time, and Display-Response Complexity on Scale Reading Efficiency (WADC technical report 58-472). Wright-Patterson Air Force Base, OH: U.S. Air Force Wright Air Development Center.

Fitts, P.M. and Jones, R.E. (1947). Psychological Aspects of Instrument Display, I: Analysis of 270 "Pilot-Error" Experiences in Reading and Interpreting Aircraft Instruments (memorandum report TSEAA-694-12A). Wright-Patterson Air Force Base, Dayton, OH: U.S. Air Force Aero Medical Laboratory.

Galer, M. and Simmonds, G. (1984). Ergonomic Aspects of Electronic Instrumentation: A Guide for Designers (Special Publication SP-576). Warrendale, PA: Society of Automotive Engineers.

Galer, M., Baines, A., and Simmonds, G. (1980). Ergonomic Aspects of Electronic Dashboard Instrumentation. in Obourne, D.J. and Levis, J.A. (eds.). Human Factors in Transport Research (volume 1). London: Academic.

Garner, W.R. (1975). Uncertainty and Structure As Psychological Concepts. Huntington, NY: Robert E. Krieger Publishing Co.

Gordon, D.A. (1981). The Legibility of Highway Guide Signs with Special Reference to Cardinal Direction Indications (report FHWA/RD-80/170). Washington, D.C.: U.S. Department of Transportation, Federal Highway Administration, June (available from NTIS as PB 82 143 868).

Graham, N.E. (1956). The Speed and Accuracy of Reading Horizontal, Vertical, and Circular Scales, Journal of Applied Psychology, 40(4), 228-232.

Green, P. (1983). What Do Drivers Say They Use Speedometers and Tachometers For? (technical report UMTRI-83-49). Ann Arbor, MI: The University of Michigan Transportation Research Institute.

Green, P. (1984). Driver Understanding of Fuel and Engine Gauges (SAE paper 840314). Warrendale, PA: Society of Automotive Engineers.

Green, P. (1988a). Human Factors and Automobile Instrument Panel Display Design (paper presented at the International Symposium on Optical Engineering and Industrial Sensing for Advance Manufacturing Technologies, Dearborn, MI). Bellingham, WA: Society of Photo-Optical Instrumentation Engineers, June.

Green, P. (1988b). Human Factors Methodology (Lecture 8). in Pew, R.W. and Green, P. (1988). Human Factors Engineering Short Course Notes. Ann Arbor, MI: The University of Michigan Chrysler Center for Continuing Engineering Education.

Green, P. (undated). Readability of Speedometers (unpublished data). Ann Arbor, MI: The University of Michigan Transportation Research Institute.

Grether, W.F. (1947). Speed and Accuracy of Dial Reading As a Function of Dial Diameter and Spacing of Scale Divisions (technical report TSEAA-694-1E) Wright Field, OH: U. S. Army Air Force Air Materiel Command, Aero Medical Laboratory.

Grether, W.F. (1948). The Design of Instrument Dials for Ease of Reading. SAE Quarterly Transactions, 2, 539-545.

Grether, W.F. and Baker, C.A. (1972). Visual Presentation of Information (chapter 2). In Van Cott, H.P. and Kinkade, R. (1972), Human Engineering Guide to Equipment Design (revised edition), Washington, D.C.: U.S. Government Printing Office.

Grether, W.F. and Connell, S.C. (1948). Psychological Factors in Check Reading of Single Instruments (technical report MCREXD-694-17A). Wright-Patterson Air Force Base, OH: U.S. Air Force Aero Medical Laboratory, September.

Grether, W.F. and Williams, A.C., Jr. (1949). Psychological Factors in Instrument Reading. II. The Accuracy of Pointer Position Interpolation As a Function of the Distance Between Scale Marks and Illumination, Journal of Applied Psychology, 33(4), 594-604.

Groth, H. and Lyman, J. (1961). A Hierarchy of "Perceptual Usefulness" of Geometrical Cues in an Overlearned Dial-Reading Task. Journal of Applied Psychology, April, 45(2), 86-90.

Heglin, H.J. (1973). NAVSHIPS Display Illumination Design Guide Section II: Human Factors (technical report NELC TD 223). San Diego, CA: Naval Electronics Laboratory Center (available from NTIS as AD 770 478).

Ishii, I. (1980). Comparison of Visual Recognition Time of Analogue and Digital Displays in Automobiles (Society of Automotive Engineers paper 800354). Warrendale, PA: Society of Automotive Engineers.

Johnsgard, K.W. (1953). Check Reading As a Function of Pointer Symmetry and Uniform Alignment. Journal of Applied Psychology. October, 37(5), 407-411.

Kantowitz, B.H. and Sorkin, R.D. (1983). Human Factors: Understanding People-System Relationships. New York: Wiley.

Kappauf, W.E. (1949). The Design and Use of Instruments (chapter 3). in National Research Council. Human Factors in Undersea Warfare. Washington, D.C.: National Research Council, Committee on Undersea Warfare.

Kappauf, W.E. (1951a). Design of Instrument Dials for Maximum Legibility: Part 5. Origin Location, Scale Break, Number Location, and Contrast Direction, (AF technical report 6366). Wright-Patterson Air Force Base, Dayton, OH: U.S. Air Force, Wright Air Development Center.

Kappauf, W.E. (1951b). Some Data on the Influence of Attempted Interpolation on the Speed and Errors of Scale Reading, (AF technical report 6530). Wright-Patterson Air Force Base, Dayton, OH: U.S. Air Force, Wright Air Development Center.

Kappauf, W.E. and Smith. W.M. (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, (AF technical report 5914, part 4). Wright-Patterson Air Force Base, Dayton, OH: U.S. Air Force, Wright Air Development Center.

Kelso, B.J. (1965). Legibility Study of Selected Scale Characteristics for Moving Tape Instruments, Human Factors, December, 7(6), 545-554.

Kurke, M.I. (1956). Evaluation of a Display Incorporating Quantitative and Check-Reading Chracteristics, Journal of Applied Psychology, 40(4), 233-236.

Lincoln, R.S. and Cahill, H.E. (1965). Detecting Out-of-Tolerance Conditons with Meter and Digital Displays. Human Factors, February, 7(1), 54-62.

McCormick, E.J. (1970). Human Factors Engineering (3rd ed.). New York: McGraw-Hill.

McFarland, R.A. (1955). Human Factors in Highway Transport Safety (paper presented at SAE Golden Anniversary Transportation Meeting, St. Louis, Missouri, October), Warrendale, PA: Society of Automotive Engineers.

- REFERENCES -

Mital, A. and Ramanan (1985). Accuracy of Check-reading Dials. in Eberts, R.E. and Eberts, E.G. (eds.) Trends in Ergonomics/Human Factors II. Amsterdam: Elsevier, pp. 105-113.

Mital, A. and Ramanan. S. (1986). Results of the Simulation of a Qualitative Information Display. Human Factors, June, 28(3), 341-346.

Morlyama, M. Kuroyama, T., and Shinkai, H. (1981). A Comprehensive Study of Digital Speedometer (Society of Automotive Engineers paper 811419). Warrendale, PA: Society of Automotive Engineers.

Murrell, K.F.H. (1962). Note on the Optimal Length for Visual Interpolation. Journal of Applied Psychology, 46(1), 41-42.

Murrell, K.F.H. (1969). Ergonomics. London: Chapman and Hall.

Murrell, K.F.H., Laurie, W.D., and McCarthy, C. (1958). The Relationship Between Dials Size, Reading Distance and Reading Accuracy, Ergonomics, February, 1(2), 182-190.

National Research Council (1949). Human Factors in Undersea Warfare. Washington, D.C.: National Research Council, Committee on Undersea Warfare.

Oatman, L.C. (1964a) Check-Reading Accuracy As a Function of an Extended-Pointer Dial Display (technical memorandum 10-64). Aberdeen Proving Ground, MD: U.S. Army Human Engineering Laboratories, July.

Oatman, L.C. (1964b) Check-Reading Accuracy As a Function of an Extended-Pointer Dial Display. Journal of Engineering Psychology, October, 3(4), 123-131.

Oatman, L.C. (1965a). Check-Reading Accuracy: Aligned Dials vs. Extended-Pointer Dials (technical memorandum 6-65). Aberdeen Proving Ground, MD: U.S. Army Human Engineering Laboratories, May.

Oatman, L.C. (1965b). Check-Reading Accuracy As a Function of Dial Alignment in an Extended-Pointer Dial Display (technical memorandum 2-65). Aberdeen Proving Ground, MD: U.S. Army Human Engineering Laboratories, February.

Oho, K. (1979). Investigation of Human-Factor Efficiency of Digital Speedometer with Electronic Display (preprint of paper presented at Society of Autmotive Engineers of Japan Meeting). See Gakujitsu Koenkai Maezurishu (Journal of Science Lecture Meeting), #791, pp. unknown.

Olson, P.L. (1963). Variables Influencing Operator Information Processing. Human Factors, April, 5(2), 109-116.

- *REFERENCES* -

Olson, P.L. (1988). personal communication.

Papaloizos, A. (1961). Some Characteristics of Instrument Measuring Dials. Ergonomics, April, 4(1), 169-182.

Pool, R. (1988). Was Newton Wrong? Science, August 12, 241, 789-790.

Poynter, D. and Czarnomski, A.J. (1988). Visual Performance with "Feature" and "Object" Displays (paper presented the International Symposium on Optical Engineering and Industrial Sensing for Advanced Manufacturing Technologies, Dearborn, MI). Bellingham, WA: Society of Photo-Optical Instrumentation Engineering, June.

Ross, S., Katchmar, L.T., and Bell, H. (1955). Multiple-Dial Check Reading: Pointer Symmetry Compared with Uniform Alignment. Journal of Applied Psychology, 39(4), 215-218.

Sabeh, R., Jorve, W.R., and Vanderplas, J.M. (1958). Shape Coding of Aircraft Instrument Zone Markings (WADC technical note 57-260). Wright-Patterson Air Force Base, OH: U.S. Air Force Wright Air Development Center.

Salvendy, G. (1987). Handbook of Human Factors. New York: Wiley.

Sanders, M.S. and McCormick, E.J. (1987). Human Factors in Engineering and Design (6th ed.). New York: McGraw-Hill.

Semple, C.A. Jr., Heapy, R.J., Conway, E.J. Jr., and Burnette, K.T. (1971). Analysis of Human Factors Data for Electronic Flight Display Systems (technical report AFFDL-TR-70-174). Wright-Patterson Air Force Base, OH: U.S. Air Force Flight Dynamics Laboratory.

Senders, V.L. (1952). The Effect of Number of Dials on Qualitative Reading of a Multiple Dial Panel (WADC technical report 52-182). Wright-Patterson Air Force Base, OH: U.S. Air Force Wright Air Development Center.

Simmonds, G.R. (1981). personal communication.

Simmonds, G.R.W. and Galer, M., and Baines, A. (1981). Ergonomics of Electronic Displays (Society of Automotive Engineers paper 810826). Warrendale, PA: Society of Automotive Engineers.

Sinclair, H.J. (1971). Digital Versus Conventional Clocks--A Review. Applied Ergonimics, September, 2(3), 178-181.

Sleight, R.B. (1948). The Effect of Instrument Dial Shape on Legiblity, Journal of Applied Psychology, April, 32(2), 170-188.

Spencer, J. (1963). Pointers for General Purpose Indicators. Ergonomics, January, 6(1), 35-49.

Terada, I., Akeyoshi, K. and Kadoo, F. (1982). Further Study on LCD Legibility Improvement for Automobiles (Society of Automotive Engineers paper 820011). Warrendale, PA: Society of Automotive Engineers.

U.S. Air Force (1987). Human Factors Engineering (AFSC Design Handbook 1-3, 3rd ed., revision 2). Wright-Patterson Air Force Base, OH: U.S. Air Force Aeronautical Systems Division.

U.S. Department of Defense (1981a). Military Handbook - Human Factors Engineering Design for Army Materiel (MIL-HDBK-759A). Washington, D.C.: U.S. Department of Defense.

U.S. Department of Defense (1981b). Military Standard - Human Engineering Design Criteria for Military Systems, Equipment, and Facilities (MIL-STD-1472C). Washington, D.C.: U.S. Department of Defense.

Van Cott, H.P. and Kinkade, R. (1972). Human Engineering Guide to Equipment Design (revised edition). Washington, D.C.: U.S. Government Printing Office.

Van Nes, F.L. (1972). Determining Temporal Differences with Analogue and Digital Time Displays. Ergonomics, 15(1), 73-79.

Vernon, M.D. (1946). Scale and Dial Reading (report APU 49), Cambridge University, England: Medical Research Council, Unit in Applied Psychology, June.

Warrick, M.J. and Grether, W.F. (1948). The Effect of Pointer Alignment on Check Reading of Engine Instrument Panels (technical report MCREXD-694-17). Wright-Patterson Air Force Base, Dayton, OH: U.S. Air Force, Air Materiel Command, Aero Medical Laboratory.

White, W.J. (1951). The Effect of Pointer Design and Pointer Alignment Position on the Speed and Accuracy of Reading Groups of Simulated Engine Instruments (AF technical report 6014). Wright-Patterson Air Force Base, Dayton, OH: U.S. Air Force Wright Air Development Center, July.

White, W.J. and Sauer, S.C. (1954). Scale Design for Reading at Low Brightness (WADC technical report 53-464). Wright-Patterson Air Force Base, OH: U.S. Air Force Wright Air Development Center.

White, W.J., Warrick, M.J., and Grether, W.F. (1953). Instrument Reading III: Check Reading of Instrument Groups. Journal of Applied Psychology, 37(4), 302-307.

Whitehurst, H.O. (1982). Screening Designs Used to Estimate the Relative Effects of Display Factors on Dial Reading, Human Factors, 24(3), 301-310.

Wokoun, W. and Chaikin, G. (1959). A Guide to Color Banding for Indicators (Meters) (technical memorandum 2-59). Aberdeen Proving Ground, MD: U.S. Army Human Engineering Laboratories, March.

Woodson, W.E. (1981). Human Factors Design Handbook. New York: McGraw-Hill.

Woodson, W.E. and Conover, D.O. (1970). Human Engineering Guide for Equipment Designers (2nd ed.). Berkeley, CA: University of California Press.

Yamaguchi, T., Kishino, T., and Dorris, J.M. (1982). The Recognition of Vacuum Flourescent Displays Under Sunlight Conditions (Society of Automotive Engineers paper 820013). Warrendale, PA: Society of Automotive Engineers.

Zeff, C. (1965). Comparison of Conventional and Digital Time Displays. Ergonomics, 8, 339-345.