LEGIBILITY OF CHILD RESISTANT BOTTLE CAPS BY TWO METHODS OF MEASUREMENT

By

Mutune Wa Gitau

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

School of Packaging

1988

ABSTRACT

LEGIBILITY OF CHILD RESISTANT BOTTLE CAPS BY TWO METHODS OF MEASUREMENT

By

Mutune Wa Gitau

As a result of difficulties encountered by consumers and physicians in reading the instructions or graphics on medical labeling, this study was undertaken to determine a method of measuring legibility. Child resistant bottle caps were used as the medium. Two instruments, a polariscope and a DMR visibility meter were used to determine legibility differences among several variables such as print color contrast, print size and visual acuity of the observer.

The comparison was determined by evaluating eight caps with two different instruction messages using twenty consumers as evaluators and a ten expert panel as a reference control. The experts' ratings were correlated with those of the consumers using the two instruments.

Results indicated that color contrast, type size and visual acuity of the observer significantly influenced legibility of the two messages on these caps. There was a high correlation between the experts' ratings and the consumers readings using the two instruments. Both instruments can, therefore, be used to evaluate legibility.

This thesis is dedicated to my wife, Wangari, and daughter, Wamaitha. Without their love, continual support, positive encouragement and patience this thesis would not have been possible. To my parents, for the foundation they gave me through effort and hard work in putting me through school.

ACKNOWLEDGMENTS

The author would like to express his appreciation and gratitude to the following individuals:

Dr. Hugh Lockhart, School of Packaging, Michigan State University, for his advice and guidance as major professor.

Dr. Susan Selke, School of Packaging, Michigan State University, for her help and direction and serving as a committee member.

Mr. Joseph Kuszai, Art Department, Michigan State University, for serving as a committee member.

Mr. Glen Pettit, School of Packaging, Michigan State University, for his ability to organize and contact the expert panel members for the legibility test.

Mr. Francis P. Tobolski, Design and Market Research Laboratory (DMR), Carol Stream, Illinois, for kindly allowing us to use DMR visibility meter for this study.

Dr. John Gill, Animal Science Department, Michigan State University, for his guidance in experiment design and statistical analysis methods.

Mr. Joshua G. Bagaka, Urban Studies Department, Michigan State University, for his assistance in analyzing the data. His countless hours of this work and genuine concern will always be remembered.

Mr. Raymond Cheung, School of Packaging, Michigan State University, for his assistance in analyzing the data, especially during a time of need.

Finally, special thanks to the students, staff members of the School of Packaging, Michigan State University, and the expert panel members who spared their time and agreed to participate in this test. Last, but not least, U.S.A.I.D. for the full scholarship granted for this program.

TABLE OF CONTENTS

Table of Contents (Continued) Page

LIST OF TABLES

List of Tables (Continued) Page

Table

LIST OF ILLUSTRATIONS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The principal functions of a consumer package, briefly stated, are: 1) to add utility or convenience to the product, 2) protection of the product against all the hazards and physical loss which may occur during the distribution cycle, 3) protection for people against dangerous or hazardous materials, and 4) communication. While the first three functions essentially deal with the technical aspects of packaging or its functional performance, communication plays the major role of informing the consumer about the product and in some cases about the package, for example opening and closing instructions. Of growing importance is the improvement of graphics for marketing, and the emphasis on health and safety for the consumer through voluntary and government standards and regulations.

Communication on a package mainly appears as written information or markings, e.g. symbols, signs, logos and pictograms, which may have different interpretations depending on the package and its targeted users. Some of this form of communication may be required by regulations.

All the printed informational material that accompanies the product is referred to as labeling, while the label is that "labeling" that is affixed to the immediate container (25). It is in this area of communication on the immediate container where legibility of the written information is critical. This is so because for written information to be communicated efficiently, it has to be legible so that the message can be read (and interpreted) correctly and absorbed in the shortest time possible. The label on a primary package that goes to the point of use is the most important in this case. Factors such as typography, quality of print, color, substrate texture, layout and the shape of the package all convey the message or communication (14, 15).

Legibility as defined by Webster's Unabridged Dictionary is "the quality or state of being legible, . . . the influence of typeface on the legibility of print" (26). Miles. A. Tinker also defines legibility as concerned "with perceiving letters, words or symbols; or the ease and speed of reading continuous textual material" (22). Legibility therefore deals with coordination of typographical factors, inherent in letters, symbols and copy material, which affect the ease and speed of reading and can only be defined in terms of a specific method of approach to the study of the problem.

The serious nature of label or package legibility has been newly recognized in two major areas:

(i) The labeling of hazardous materials - this is concerned with the legibility (readability) of use instructions and warnings on the label. The labeling requirements for hazardous materials are regulated by three bodies: the US Department of Transportation (DOT) under the Title 49 of the code of Federal Regulations (CFR) in parts 171-179, the Consumer Product Safety Commission (CPSC) Title 16 CFR part 1700 of the Poison Prevention Packaging Act, and the Environmental Protection Agency (EPA) under regulation 40 CFR 162.10 and in section 162.16 regarding Child Resistant Packaging (CRP).

(ii) Medical labeling - Fast, accurate identification of drugs is critical in a crisis. Of particular concern recently was the problem of medication errors in hospitals by anesthesiologists during surgery which occurred due to incorrect identification of injectable drug products (4). This led to the formation of Subcommittee D10.34 on Identification of Pharmaceutical Drug Pruduct Containers of the American Society for Testing and Materials (ASTM). This subcommittee of ASTM Committee D-10 on Packaging has addressed the issue of legibility in the "Standard Specification for Labels for Small-Volume (Less Than 100 Ml) Parenteral Drug Containers, D4267-83" (1). This includes label legibility requirements, legibility tests, type size requirements and requirements dealing with the orientation of the message on the container. The specification requires a minimum type size of not less than legible 6-point type for the proprietary or established name of the drug and the numerals indicating the amount of drug per unit.

The legibility test in the specification requires that the proprietary and established drug names, and the amount of drug per unit, be legible in a light of 20 foot candles at a distance of 19.7 inches (500 mm) to a person with 20/20 unaided or corrected vision. This test was chosen partly because many drug labels were difficult to read and partly because of the dim light found in hospitals. This specification was expected to alleviate the previous problem of identification in the hospitals.

The other area of increasing interest is the legibility of child-resistant bottle caps. These caps are designed to protect small children from a dangerous product. However, if the package labeling does not inform the adult how to use the package, the protective function may be lost. So, for lack of effective communication, the protective function becomes inoperative. Child-resistant packaging caps (CRP) in the market today need to be legible but many consumers have reported difficulty reading the white-on-white print on many bottle caps. This problem is particularly prevalent among the elderly population. In its report on Innovative Child Resistant Packaging Systems (16), the School of Packaging reported that in the past, messages on child resistant caps were often printed in colored ink. From talks with caps producers and user companies, it was learned that this practice was diminished by manufacturers because of the cost of operation and the poor quality of the printing. However, industry members interviewed agreed that the caps instructions and package graphics could be printed economically by using non-contact printing methods, where the printing ink would cost about one cent per pint. This would produce an acceptable printing that would remain neat and legible, avoiding the chipping or rubbing off the package during shipment or repeated use, which has been a problem.

The objective of this research was to devise a method or system of measuring the legibility of labeling. Child resistant bottle caps were chosen as the study medium. The labeling on these caps was the main visual stimulus in this test. To evaluate the legibility and to compare the methods, the study involved two groups. One was an expert panel (professional people working in graphics trades). The other group was consumers in two age subgroups: 18-25 years and 30-54 years. These two groups measured the legibility of child resistant caps by the use of two mechanical devices which were selected for this study--the polariscope and the Design & Market Research Laboratory (DMR) visibility meter.

The results of the research are expected to help in answering the following questions.

1. Do the instrument readings give the same result as the experts?

2. How do the two instruments compare? The results will help determine whether there is any significant differences between the two devices used.
3. Effect of contrast and type size on the legibility of the message.
4. Effect of visual acuity of the observer on the legibility of the message.

The message content and understanding of meaning were not measured, only the ability to detect and read the words under a controlled amount of illumination.

CHAPTER 2

LITERATURE REVIEW

A major goal of this work has been to establish a convenient and reliable method for determining the legibility of package label copy. Some methods have been used previously and described in the literature. Seven such methods are described in the following pages.

2.1. Distance Method

This method has found considerable use in studying legibility, particularly visibility or perceptibility at a distance (23). The apparatus consists of a wooden rail about three meters long placed before the subject, slanted downward at about a 15 degree angle. There is a headrest at the upper end of the rail. The stimulus material (letters, words etc.) is placed in a small, well illuminated car which can be moved any desired distance along the rail. The stimulus material is placed near the far end of the rail and the car is moved towards the subject by steps of about 10 centimeters at a time, thereby allowing the subject to read as much of the material as possible at each step. A modification of this method was developed to study the legibility of large print for advertisements or highway signs. Here the "stimulus material is placed in a fixed position and the subject walks slowly toward it until he just reads it correctly" (23), and the distance is recorded.

This method measures the distance from the eyes at which print can be perceived accurately. It has been found useful in studies of relative legibility of letters of the alphabet and digits, of specific letters in different type faces, and the effects of brightness-contrast between print and paper surface. However, Tinker states that this method

> "has little or no validity for investigating such typographical factors as leading, line width, determination of optimal type size for reading continuous text, or optimal type faces." (23)

However, it has been used to measure the legibility of print for road signs, bill boards and poster advertising, which involves the use of large print. Apart from this, the application of results obtained by this method to the legibility of print has not been ascertained.

2.2. The Tachistoscope (T-Scope) or The Short-Exposure Method

M. A. Tinker describes the short-exposure method as the method for measuring "legibility by determining the speed of accurately perceiving printed symbols." (23). The apparatus is called a tachistoscope. This is a mechanical device used to evaluate surface designs of a package. Leslie Barton (3) defines the tachistoscope as an instrument which measures legibility, recognition, attention and memory value which may be correlated with shelf visibility. Barton says that this is a useful testing device where individual word legibility is in question. Donald Morich (13) describes the tachistoscope as one used to measure the impact of salient visual elements of a package and recommends it for most marketing-packaging evaluations, particularly as an impact measure. This device is comprised of a shutter that, for a desired fraction of a second, will allow a subject to view a test sample. The device precisely controls the amount of time a stimulus (i.e. a package) is exposed to the subject or respondent, thus ensuring equal exposure time.

Barton identifies three types of tachistoscopes used in Packaging--slide projection, mechanical and light flash (3). Slide projection is the most common type of tachistoscope in use today. It is a still projector with a shutter similar to that used in a camera. The slide projector flashes the image on a screen and the duration is controlled by changing the setting on the shutter. The triggering device is the most important physical consideration. The tachistoscope can be mounted to a normal slide projector. Advantages include group viewing, flexibility as to physical sample size, and the increased attention of the subject when showing a lighted area in a dark room. One disadvantage, though, is not seeing an actual package. The actual testing should expose each subject to a preset number of exposures starting at a speed of 1/150 seconds and graduate down to speeds around 2 seconds. After each exposure the respondent reports what he saw according to a pre-established question guideline. People do become more proficient as the test continues, so the order of samples should be uniformly varied and a dummy sample used initially for warm-up. Lee Swope (17) emphasizes the use of a test group with varying eyesight as this would tend to simulate the real market place. He says the test reveals the impact measure on brand-name recognition, product description play back, recognition of symbols/logos and other salient graphics. This can then be used to compare design alternatives such as shape, color and type. Swope goes on to say that this technique is ideal in comparing test package alternatives as it produces consistent results in a test-retest situation and is able to discriminate among test alternatives. Similarly, if the test packages are capable of producing different levels of impact, this testing procedure will measure the difference.

Another form of tachistoscope in use is similar to a revolving blackboard on which a multitude of objects are placed and the viewer's response recorded. The viewing period is usually in seconds rather than fractions of seconds. In this case, there is a variation in lighting, and objects (packages) are viewed, rather than images as in the case of slide projection method of tachistoscope. The third type of tachistoscope described by Barton uses a flash of light on an object (firing a flash of light of accurate duration by electrical impulses on a magnetic tape). Flash can also be originated by construction of a tachistoscope using a high speed photolight which can achieve exposures of one millionth of a second. The major advantage is the accuracy achieved at short durations; objects (packages) are viewed as opposed to slide projector images (3).

The tachistoscope is useful as an impact measure, according to Barton, but does not reveal the legibility of the graphics on a package particularly as they pertain to textual information on the packages (3). This instrument is important for such features as brand name recognition, logos and symbols, which play a major role in marketing of a product. In areas where detailed information is important to a consumer or a professional, such as in medical or pharmaceutical labels/packages, the tachistoscope cannot be used. This is because the exposure time is very limited, time here is a controlled factor, and the level of illumination is changing greatly in a short time. Alf Nelson, the inventor of the polariscope for package graphics evaluation states that

> "legibility cannot be accurately measured during changes of illumination, because the time lag necessitated by the pupillary adjustments of the eye under varying levels of illumination creates false reading results" (18).

M. A. Tinker reported that

> "the relationship between tachistoscopic reading and ordinary reading of continuous textual material is so small that conclusions concerning legibility of continuous printed text from tachistoscopic results must be made with caution" (23).

In summary, this method is only useful for measuring the impact and recognition of letters and digits, specific letters in different type faces and the effect of variation in brightness contrast between print and paper.

2.3. The Focal Variator Method

This method consists of a system of lenses which are related to each other such that a visual stimulus (print) may be projected upon a ground glass screen in any degree of clearness from an unrecognizable blur to clear definition of focus and thus the degree of clearness can be measured (23). To facilitate motion, the lenses are interconnected in such a way that they travel in opposite directions. As one lens increases the size of the image, the other reduces its size. This results in a blurred image on the ground glass having the same size as the clear image would be.

No image is present on the ground glass at the start of the test, but as the subject turns the wheel of the apparatus, the image appears and comes into focus. The respondent or subject reports as soon as he perceives a meaningful visual form. Scale value on the apparatus is also recorded, which indicates the degree of clearness of the visual task at every step. A zero scale value indicates maximum clearness.

This method may be used to determine legibility by measuring how far a letter can be thrown out of focus and still be recognized. The technique is however limited to the investigation of the relative legibility of letters of the alphabet, of digits and of specific letters in different type faces. The apparatus has been reported by Tinker (23) to yield precise measurements of relative legibility of the above which should have high validity.

2.4. Rate of Involuntary Blinking Method

Luckiesh and Moss promoted blink rate as a measure of readability (legibility) (23). The studies were concerned with ease of seeing and the effects of varying typographical factors. The rate of involuntary blinking during reading is assumed to be inversely proportional to ease of seeing. This means that the easier the visual task, the fewer the blinks for a set period of reading. Therefore any visual task (i.e. typography) which is read with greater ease should produce fewer blinks. Similarly, the rate of involuntary blinking decreases as the level of illumination increases, while it is greater for the lower levels of illumination.

In studying typographical factors, the eye blinks are counted by direct observation for a period of five minutes while the subject is reading continuous text printed in each set up (i.e. 6 pt or 12 pt type etc.), while the illumination is held constant for all the tests (23). The blinks are counted and recorded by the experimenter seated to the side and slightly behind the subject. Tinker reports that, to study the effects of illumination level on reading, the blinks are recorded during the first and last five minutes of approximately an hour of continuous reading under each light intensity level (23).

Tinker evaluated the reliability and validity of legibility measurement by this technique during reading by using 74 and 64 subjects in two experiments. For adjacent five minute periods, the reliability coefficients were fairly high, .74 for the 74 subjects and .92 for the 64 subjects. For periods of 10 to 15 minutes a similar trend was observed. When an interval of 20 minutes of reading was substituted, the reliability dropped to .49 to .56 which was marginal and indicated inconsistency of the blink rates (23). It is suggested that to minimize the effects of individual differences, sizable groups of subjects should be used in order to have a higher reliability especially where group comparisons are concerned.

The validity of this method in measuring legibility has not been clearly ascertained even though the results comparing the reading of 12 pt Book type and 7 pt the Newsprint by Tinker appear convincing. The data indicated that the Newsprint was easier to read than the Book type where the scores were recorded from 60 readers whose frequency of blinking was recorded at 5 minutes intervals for a total of 10 minutes for both typefaces. It is further reported that none of the results from Luckiesh's laboratory were evaluated statistically (23). This brings into question the use of the rate of blinking as a criterion of ease of seeing or legibility since it is doubtful that the frequency of reflex blinking can be accepted as a valid measure because this can also be affected by other psychological and physiological factors which might affect the results.

2.5. The Eye Movement Test

The eye movement test indicates where the eye is travelling on a package, and the sequence in which the different graphic elements are being observed. This test is important as an analysis of eye movement patterns which can determine how well the overall graphics layout and typography attract or hold the eye, while it yields information on why one typographical arrangement is more or less legible than another. Tinker reports that "eye-movement records tend to be reliable and valid measures of legibility of print" (23).

There are different techniques which are used to record eye movements in reading. One form is that of a single movie camera fitted with a special close-up lens which takes pictures of the eye and transmits them to a computer controlled unit which bounces a beam of infrared light off the retina, thus tracking the eye movement physically as the subject views the test package (17). Each movement of the eyes produces a change in the beam of reflected light so that horizontal and vertical movements are recorded on the film. Tinker mentions that the eye movements are readily distinguished from the head movements on the film record (23). This is also referred to as the corneal reflection method. The test shows what the subjects observed in a given amount of time and in a certain sequence. The less important graphic elements that may interfere with those elements which are essential to package communication are highlighted by this test, and this may result in corrective measures for redesigning the package. The major disadvantage of this test is that it does not show the quality of the graphic elements observed in detail. This is where a tachistoscopic test is helpful to measure noticeability of individual graphic elements.

The other form of eye movement test commonly used is the electrical method, which depends upon an electric potential difference which exists between the cornea and the retina of living eyes (23). The movements are recorded by placing electrodes on the skin near the eye and attaching the electrodes to an amplifying system and an ink-writing oscillograph or a photographic device. This method has gained more credibility because it is much more flexible than the corneal reflection method especially because the subjects (or reader's) head is not held rigidly in one position and records for longer periods of reading can be obtained.

Results obtained by Tinker for short paragraphs found reliabilities from .66 to .89; for reading 20 to 25 lines; the mean reliability was .81 (23). He therefore reported that those measures had satisfactory reliability for studying legibility of print and all eye-movement measures provided valuable supplementary information in investigating legibility. However, it is Tinker's opinion that these measures and results can only be correlated to reading performance since they do not show the degree of legibility of individual graphic elements but largely deals with a subject's perception and reading ability.

2.6. Polariscope (P-scope)

Leslie Barton describes the polariscope as an instrument which can be used to measure the various degrees of visibility or legibility of the different components of a surface design (3). The one used by Barton for his legibility study was an adaptation of the polariscope based on the original Alf Nelson (18) invention which was referred to in the literature as "Polarascope" visual tester. The polariscope used for the study reported here was designed by General Radio Company (5).

Barton's description of the operation of the polariscope is summarized as follows: Polariscope is based on the properties of polarized light where two light polarizing screens are used. The two screens are the polarizer and the analyzer. Light is polarized by the first screen, the polarizer, and is passed through the analyzer. When the screen polarizing planes are parallel to each other, the maximum amount of light is allowed to pass through and when the planes are perpendicular the opposite is true as all light passing through the polarizer would be absorbed by the analyzer. Both of these results depend upon the relative orientation of the screens. Each plane polarizer carries a degree scale calibrated from zero to plus and minus ninety degrees for determining the degrees of rotation (3). The normal light source furnished with the polariscope is an incandescent lamp contained in a metal housing with a diffuser fitted to it. This light source is used when light transmitted through a material is used to evaluate stress patterns.

The polariscope has been widely used for dynamic and static stress analysis in the field of mechanical research. Of importance in this area is photo-elastic stress analysis, an experimental method of determining where and to what degree stress concentrations will exist in a plane section under known loading conditions. The method has been applied successfully to design problems ranging from bolt heads to battleships (5). More important to a packager is the use of the polariscope in evaluating seal integrity in pouches and strains in glass by observing color differences of transmitted lights which are related to stress concentrations. A further use in packaging has been the adaptation of the polariscope to evaluation of the visibility, or legibility of label elements, as described by Barton (3).

Lockhart and Michel (8) describe the adaptation of the General Radio equipment in summary as follows: The label element is placed between the diffuser and the polarizer screen on a platform where it is slanted to an angle for easy viewing. A source of illumination is directed at the label element at this angle from a lighting source. Once the light falls on the label, it is then reflected through the polarizing screen and on to the analyzer screen and this light is viewed by the subject rotating the analyzer screen.

In comparing the tachistoscope with the polariscope, Barton stated that the tachistoscope tests revealed the attention-getting characteristics of a package as observed by a shopper when he gets a glimpse of the package, while the polariscopic tests revealed the order in which the subject would perceive the surface design after his attention was brought to it and he stopped and studied it for a few seconds (3). Barton reports that specific information is not seen as a whole through the polariscope as is the case with tachistoscope; it is usually seen in parts.

Alf Nelson, the inventor of the polariscope application claimed that eyesight differences would not affect the results because all surface designs were rated comparatively by each individual (18). However, tests conducted by Lockhart and Michel, of the School of Packaging, Michigan State University (8) showed that people with 20/30 vision needed considerably more light to see a specified message than those with 20/20.

Using the polariscope, Lockhart and Michel tested 48 subjects, 81% of whom had 20/20 vision and the rest with 20/30 (8). They were tested on two package forms, a bottle with a printed label and an ampoule with message silk screened on the surface. For both packages, as the type size got smaller, the degrees of rotation increased--showing that more light was required to read the message. In a second test by Lockhart and Michel, there were eight variations of black on white printed labels comprised of two messages, two type styles (Garamond and Helios) and three type sizes of 5, 6, and 10 points. After developing a split plot experiment utilizing analysis of variance, the results showed the effect of the difference of the two label messages with the two different type styles--with the Garamond being harder to read than Helios. Similarly, the subjects with 20/30 vision had greater difficulty reading the labels than did the subjects with 20/20 vision.

The outcome of this experiment showed that the polariscope can measure differences in legibility between type sizes and type styles. The authors say that this method could also be adapted to pass-fail testing as well as comparative testing on label design as follows: since the experiment involved medical labels, it was the two authors' belief that

> "the legibility test should take into account that the medical practitioner is expecting a particular message on the label; he is comparing the message on the label with the message in his mind--not seeking to decode a totally unfamiliar message." (8)

One of the most important advantages of the polariscope is its ability to determine which design elements will be most legible to the observer and to aid in finding design criteria for choosing the package surface design. The only major disadvantage pointed out by Barton is that, "the polariscopic tests indicate nothing as to how the design will perform among others as in a store display" (3). This is true because polariscopic tests involve single packages which are normally viewed in parts and in isolation.

2.7. Luckiesh-Moss visibility meter and Design & Market Research Laboratory (DMR) visibility meter

The Luckiesh-Moss visibility meter was developed by Luckiesh and Moss as a means of measuring visibility under controlled conditions of illumination, thus obtaining a foundation of knowledge upon which to specify light and lighting for specific visual tasks (10). This meter is a binocular device that has two colorless filters in parallel, one for each eye, which are optical circular gradients. Each filter varies from almost clear to very dark, and when rotated simultaneously in front of the eyes, can alter the brightness--contrast of the object whose visibility is to be measured. The amount of rotation is marked by a numerical circular scale associated with the filter. The instrument is held in approximately the same position that eye glasses are worn, and with the use of a finger, a knob is slowly turned until the visual object or its fine detail is legible to the observer, at which point the reading is recorded (10).

The meter scale is called relative visibility (RV) and it has a range of 1 unit to 20 units. A second scale on this instrument is that of relative foot candles which ranges from 1 to 1000. By definition, the scale term "relative" visibility can be "absolute" visibility, while the scale of relative foot candles is related to the amount of light required to achieve a given visibility (12). The foot-candle scale is not of interest for this investigation, but the relative visibility scale is.

In designing the scale and calibrating the visibility meter, some assumptions were made by Luckiesh and Moss. They pointed out that,

> "critical visual tasks are performed either at distances within arm's reach, as in reading, or at much greater distances, as in driving an automobile. The ophthalmologist usually assumes that distances of about 14 inches and 20 feet, respectively, are representative of these two cardinal distances." (12)

Therefore, the visibility scale can be calibrated for either near-vision or distant vision, with respect to the detection of the presence of an object and its critical resolution. This scale is calibrated in terms of the threshold size of Luckiesh and Moss standard parallel-bar test. This test uses a standardized object which consists of two parallel bars of certain width and spaced a distance apart equal to the width. The dimension of the bars and space are related to the viewing distance in such a way that the space/bar dimensions and the viewing distance establish a known visual angle. The threshold size is the size that is just visible under 10 foot candles illumination by a person with normal vision (12). For the scale design, and calibration object dimensions and viewing distances, combinations were chosen to given visual angles ranging from 1 minute subtended angle to 20 minutes subtended angle. Subjects of normal vision then viewed the objects through the meter and rotated the filters until the objects were just visible. The scale was them marked in terms of the visual angles 1 minute to 20 minutes.

Tinker points out that since the meter is calibrated in terms of threshold size, a visibility scale value of "1" indicates the standard test-object (parallel bars) whose critical detail subtends a visual angle of one minute. Similarly, a scale value of "2," two minutes, etc. (23). Thus, the scale values of relative visibility are directly proportional to the visual size of the standard test-objects under threshold conditions.

When the instrument is calibrated and ready for use under standardized conditions, it is possible to measure the visibility of any object. A standard condition for viewing the objects is provided by the simple method of having the field surrounding the object either black or of constant brightness.

Among the studies conducted using this meter is that of relative visibility of type faces, type forms, size of type, and effects of variation in brightness contrast between print and paper. Further studies by Luckiesh and Moss explore the relationship of illumination intensity to type size. It was reported that,

> "about 3 1/2 times as much illumination was required to make 6 point Bodoni type as visible as 12 point. Deficiencies in type sizes between 6

> and 12 point can be compensated for by increases in illumination." (10)

The same study found out that the visibility of 3-point-type was close to "1," which indicated that this was about the smallest type readable by persons with average normal vision under an illumination of 10 foot candles. One drawback of these measurements as argued by Tinker (23) is that one does not know what scale value corresponds to optimal legibility when investigating size of type. The author continues to say that for studying the effects of variation in type face, in leading or in line width, visibility measures are of little value. Nevertheless, situations where visibility scores are useful include studies of the effects of brightness contrast between print and paper, and the relative legibility of letters of the alphabet, digits and other isolated symbols.

The Design and Market Research visibility meter (DMR) which was used for this study was designed by Francis P. Tobolski (24). The meter was designed using the same principles used in the original Luckiesh-Moss visibility meter. The difference from the original Luckiesh-Moss meter is that the DMR meter was redesigned for the use of measuring legibility/visibility and therefore the scale for relative foot candles has been eliminated and in its place is the scale of relative visibility. The scale begins at 1 and ends at 20 just like the original meter. Another modification of this meter is the provision of a handle, which a subject or viewer can use while putting the meter in front of the eyes and with the use of the other hand (or fingers), can rotate the disks comfortably. This modification clearly improves the usability of this device because of the efficient handling device.

One of the precautions to be taken while using a visibility meter, according to Luckiesh (10), concerns extraneous light, especially from sources and areas in the visual field surrounding the object or field. If allowed to enter the windows of this meter, it can add to the veil of haze, thereby affecting the visibility measurement. This can be prevented by either screening the visibility meter from extraneous light or by maintaining a constant level of illumination. For the purpose of this study, the latter was chosen with illumination maintained at an intensity of 50 foot candles.

CHAPTER 3

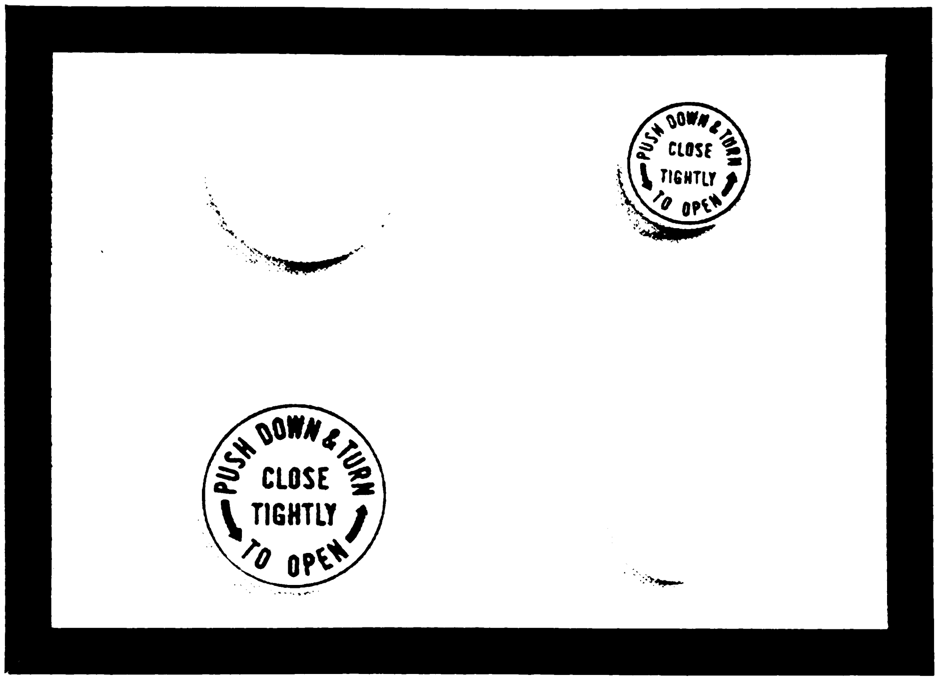
MATERIALS AND APPARATUS

3.1. Materials

Testing for the legibility of messages on child resistant caps was performed on two types of cap system of similar (plastic) material, but with different message content. The message, which consisted of opening and closing instructions, was emboss printed on the top surface of the caps. The emboss printed message was white on a white background and all the caps were white in color. For the purpose of this study, the lettering of half of the caps used for this study were colored black in order to compare the effect of contrast differences on legibility. For the purposes of simplicity in the presentation and discussion of results, the following message categories and designations will be used.

Message I - In this category were 2 paired (four) caps of two sizes (33 and 22mm continuous thread caps) with the following instructions: "Push Down & Turn, To Open," including two directional arrows; and "Close Tightly." (See illustration No. 1) This message appeared in bold upper case sans serif type style on both pairs of the test caps.

ILLUSTRATION NO. 1
MESSAGE 1 CRP CAPS



THE CAPS ON THE LEFT ARE IN 12PT TYPE SIZE WHILE THOSE ON THE RIGHT ARE IN 8PT.

The message type size was 12 pts on the large pair and 8 pts on the smaller pair. Although all the test caps were originally supplied emboss printed white-on-white, the lettering on one cap of each pair was colored black in order to create a color contrast for testing purposes. These caps were supplied by Kerr Glass Manufacturing Corporation, Lancaster, Penn.

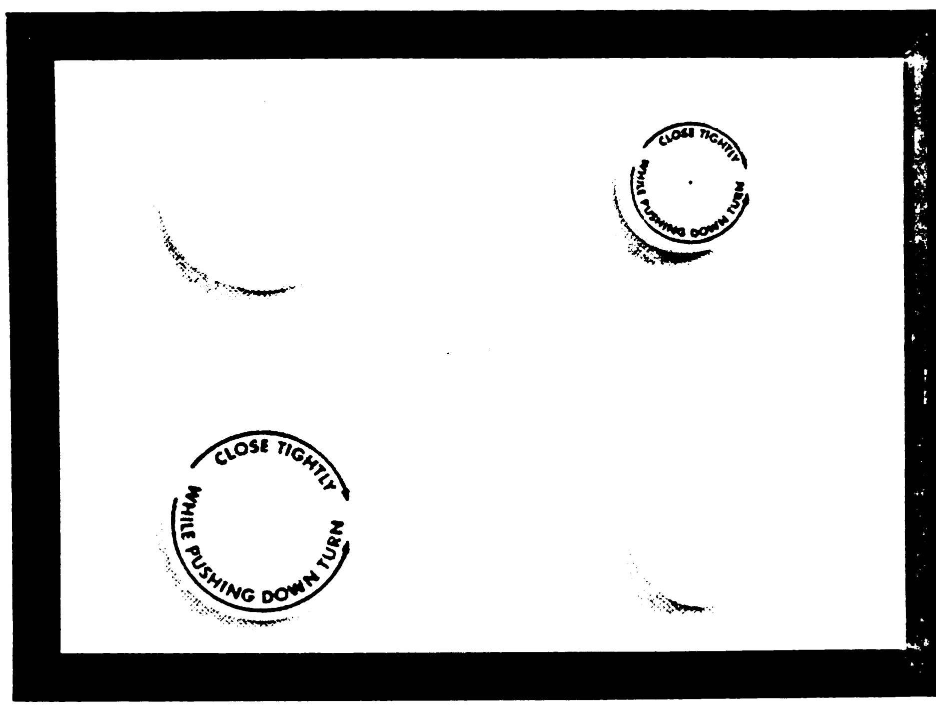
Message II - Same pairs and cap sizes as above, but with the following message instructions: "Close Tightly," and "While Pushing Down Turn" including two directional arrows. (See illustration No. 2) The message on the larger pair was emboss printed in bold upper case sans serif type style in 8 pts while the smaller pair appeared in 6 pts. These caps were supplied by Owens Illinois, Brookville, Penn.

3.2. Major Apparatus

Two types of mechanical devices were used to determine the legibility of messages on the child resistant caps. The main purpose of using the two devices was to compare the legibility measurements obtained with these devices and to correlate the results of each with the expert panel's evaluation as the control, with the ultimate goal of recommending the most suitable method of measuring legibility. They are the following:

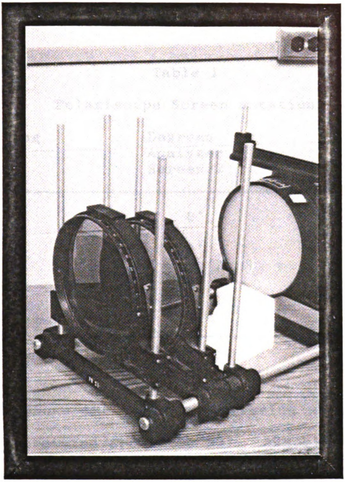
(i) Polariscope - Polariscope Type 1534-A, manufactured by General Radio Company, Cambridge 39, Massachusetts, was used (See illustration No. 3) The major components were the

ILLUSTRATION NO. 2
MESSAGE 2 CRP CAPS



THE CAPS ON THE LEFT ARE IN 8PT TYPE SIZE WHILE THOSE ON THE RIGHT ARE IN 6PT.

ILLUSTRATION NO. 3
POLARISCOPE



THE POLARIZING SCREEN AND THE ANALYZER WITH THEIR SCALES ARE IN THE FOREGROUND. THE CRP CAP IS LOCATED ON THE WHITE BLOCK BETWEEN THE POLARIZING SCREEN AND THE DIFFUSER (IN THE BACK GROUND).

Polarizer and Analyzer Type 1534-PI, both with a scale of degrees of rotation from 0 to 90 degrees. The following table shows the Polariscope screen rotation (9).

Table 1

Polariscope Screen Rotation

Angular Setting Polarizer Screen 1	Degrees Analyzer Screen 2	Light Conditions
90°	0°	Dark
90°	10°	Lighter
90°	45°	Still Lighter
90°	90°	Maximum Light

Table 1 shows the light transmission when the analyzer (screen 2) is rotated from 0 to 90°. Assuming the polarizer is fixed at 90° when the analyzer reaches 90°, its axis of polarization is parallel to that of the polarizer and the maximum amount of light is transmitted.

(ii) Illumination. A controlled source of illumination for the polariscope was provided by a General Electric (GE) high intensity bulb of 40 watts (120V), with white reflector.

(iii) Design and Market Research (DMR) visibility meter - The binocular device used for this study was supplied by Design & Market Research (DMR) Laboratory

Corporation, Carol Stream, Illinois, (See illustration No. 4). This instrument is based on the use of a neutral density wedge and an opaque to transparent wedge. The visibility scale runs from 1 to 20' (visual angle in minutes). The source of illumination was the same as that used with the polariscope. The following table shows the visibility meter's opaque to transparent wedge rotation, when rotated from 20 to 1.

Table 2

DMR Visibility Meter Wedge Rotation

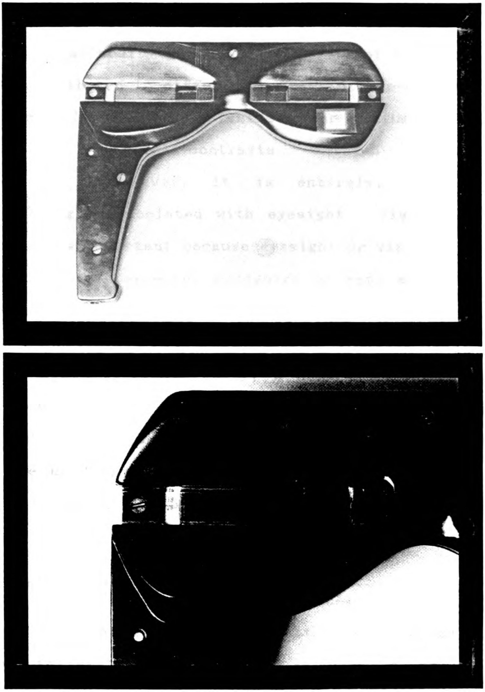
Visual Angle	Light Conditions
20	Semi-opaque
15	Lighter
10	Still Lighter
1	Maximum Light (Transparent)

3.3. Other Requirements

(i) Light Meter - This was used for the purpose of measuring the amount of illumination used in this test. The amount used for testing purposes was 50 foot candles. This level of illumination was arrived at as a result of a general survey conducted in the offices of the School of Packaging and one department Store in East Lansing, where

ILLUSTRATION NO. 4

DMR VISIBILITY METER AND A CLOSE-UP SHOWING THE SCALE AND THUMB WHEEL ON THE RIGHT



illumination was recorded at an average of 50 foot candles. The light meter used had a range of 0-250 foot candles and was manufactured by General Electric Corporation, Type DO-78, Model J-55. No. F-2, (See Illustration No. 5).

(ii) Near Point Visual Acuity Card - Visual acuity is commonly defined as the ability to distinguish fine details (11). Visual acuity may be high or low for a person, which can be attributed to different degrees of eye-defectiveness, different levels of brightness or illumination, and different brightness-contrasts between object and background. However, it is entirely, or at least overwhelmingly associated with eyesight. Visual acuity in this test is important because eyesight or vision influences the ability to perceive, recognize or read a given visual message such as print under a controlled level of lighting or illumination.

To measure visual acuity, a near point visual acuity card from Dow Corning Ophthalmics Inc., Norfolk, Virginia, was used, (See illustration No. 6). The instruction was to hold the card in good light 16 inches from the eyes. The Dow Card follows the Snellen System (11) of rating vision, where:

$$\text{Visual acuity} = \frac{d}{D} = \frac{\text{actual distance}}{\text{normal distance}}$$

"d" is the distance in feet at which a given line of letters is barely recognizable by any subject, while "D" is the distance in feet at which the same line of letters is

ILLUSTRATION NO. 5
LIGHT METER

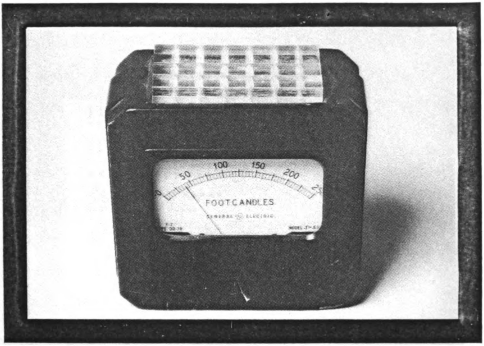
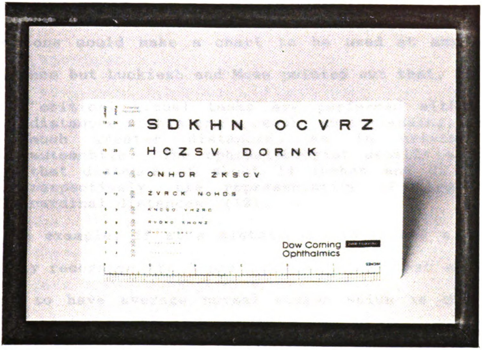


ILLUSTRATION NO. 6
NEAR POINT VISUAL ACUITY CARD (DOW CARD)



barely recognizable by a person with normal vision. "d" is always the distance at which the test-chart is actually viewed. The Dow Card has Distance Equivalent ranging from 20/200 (in 26 pts of assorted letters of the alphabet in upper case) to 20/20 (in 3 pts). The Dow Card is devised to be used at a distance of 16 inches. Luckiesh (11) emphasizes that one could make a chart to be used at any specified distance but Luckiesh and Moss pointed out that,

> "critical visual tasks are performed either at distances within arm's reach, as in reading, or at much greater distances, as in driving an automobile. The ophthalmologist usually assumes that distance of about 14 inches and 20 feet, respectively, are representative of these two cardinal distances" (12).

As an example, if at a distance of 16 inches a person can barely recognize the 3 pts letters in the test chart, he is said to have average normal vision which is expressed as 20/20 vision. If one can barely distinguish the 26 pts letters (first line on the chart), his vision is 20/200. This means he is just able to recognize at 20 feet what a person with normal vision can recognize at 200 feet. Therefore threshold size for him is 10 times normal and his visual acuity is rated as 10% of average normal (11).

CHAPTER 4

TEST PROCEDURE

4.1. Introduction

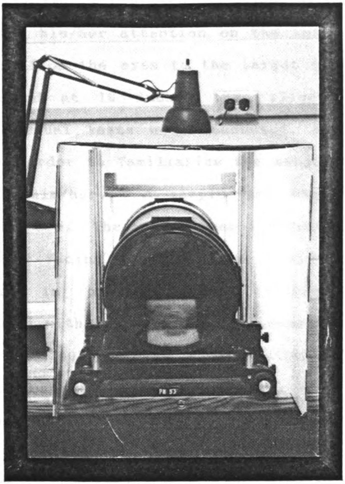
Tests were conducted to determine the legibility of white-on-white and black-on-white typographic messages on child-resistant caps. Two devices were used in conducting this test--the polariscope and the DMR visibility meter. Two other objectives were to determine if these two devices gave the same information on legibility and whether or not they gave the same result as would expert evaluators. In these tests, two classes of consumers (based on age subgroups) evaluated eight CR closures with the polariscope and DMR visibility meter. The first age subgroup was made up of ten students from the School of Packaging, ages 18 to 25 years. The second age subgroup consisted of ten staff members from the School of Packaging ages 30 to 54 years. A third group was made up of an expert panel consisting of ten professionals in graphics trades related to legibility, and this group was used as a standard or reference control. Visual acuity of each group was measured with the Dow Card. In this test, the illumination was held constant at 50 foot candles.

4.2. Polariscopic Procedure

Ten subjects of the first age sub-group (18-25 years) were used in this test. Before the beginning of the tests, every subject read a near point visual acuity card using both eyes and wearing such vision correction as was normal for them. This visual acuity was then recorded on the test form (see sample Test Form No. 1, page 51). 70% had a 20/20 visual acuity while the rest (30%) had a 20/30 acuity. The other ten subjects tested on this device were from the 30-54 years age group, 70% of whom had 20/20 vision and 30% 20/30 vision.

The polariscope used for this test was placed on a table and covered by a hood (see illustration No. 7) This hood was a corrugated box with a circular opening on the top to allow for a controlled light source, and an opening at one end so that the viewer could rotate the analyzer screen of the polariscope to view the object. The inside of the corrugated box was lined with chrome-coated paper board to help prevent absorption of light by the corrugated board, thereby maximizing the illumination intensity. The lighting or illumination was held constant throughout the test at 50 foot candles. Only the test source of lighting was used for this test, and therefore the rest of the lights in the laboratory were turned off during the duration of this test. The test source of lighting was a General Electric 40 watt bulb which was mounted in a lamp with white reflector.

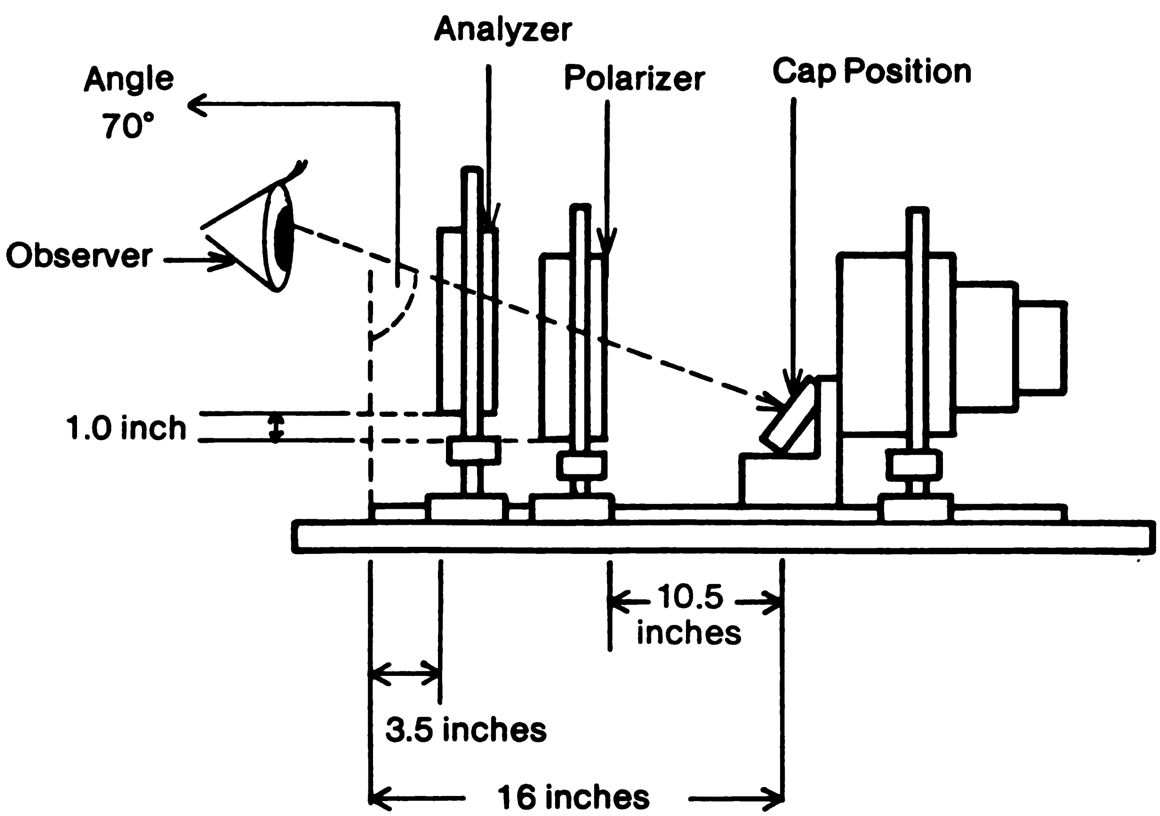
ILLUSTRATION NO. 7
POLARISCOPE COVERED BY A HOOD (CORRUGATED BOX)



At the beginning of the test, the polariscope screens were closed so that the closures could not be seen. This means that the analyzer (screen 2) was at zero degrees while the polarizer (screen 1) was at ninety degrees. In the test room, the subject was seated in front of the polariscope which focused his/her attention on the instrument screens. The distance from the eyes to the target (or the test cap) was maintained at 16 inches (See illustration No. 8). Before the actual tests were conducted, a trial test was performed in order to familiarize the subject with what was expected of him/her. A sample cap was placed in the illuminated space. The subject was instructed to rotate the first screen facing him/her (the analyzer) slowly and consistently in one direction (either clockwise or anticlockwise) with his hands until a message on the cap was legible. The subject was then instructed to stop the rotation and report what he/she could read. Subjects were not instructed to read in any special order, but were encouraged to indicate when they could read the whole message. The cap message was not revealed to the subject.

Once the subject read the message, the numerical reading on the scale attached to the outer diameter of the analyzer (screen 2) was recorded by the test monitor. This scale started from zero (with no visibility) to 90 degrees, the maximum allowable light transmission. After familiarization with the evaluation procedure, the sample cap was replaced by the test cap and the above procedure was

ILLUSTRATION NO. 8
A DIAGRAM SHOWING THE POSITION OF THE OBSERVER IN RELATION TO THE P-SCOPE AND THE TEST CAP.



repeated. The reading was a direct mechanical reading as no electrical or other devices were used to determine the reading. Each result was recorded on a test form (see sample Test Form No. 1, page 51).

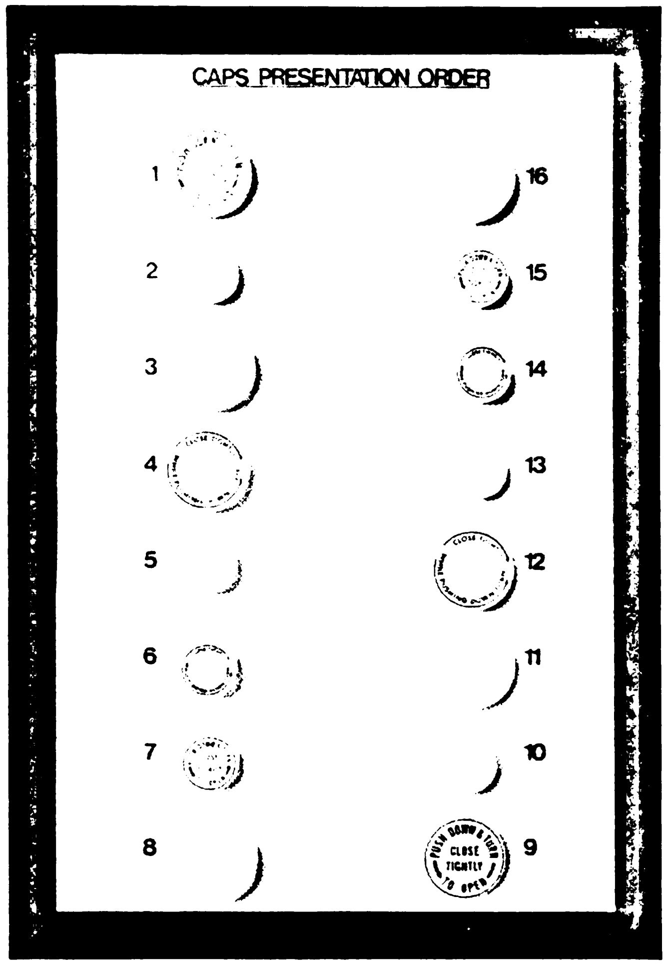
The eight caps (see illustration No. 9) consisting of four caps with message I and the other four with message II were presented to the subject in a fixed order (see Table 3) and each cap was viewed once. One set of the caps with message I was mixed with the other set of message II and numbers 1 to 8 were assigned to them. The caps were then viewed one at a time in the same sequence by all the test subjects.

Table 3

Cap Order of Presentation to the Consumers on Both Polariscope and DMR Visibility Meter

Cap No.	Color	Size	Message Type
1	Black-on-white	12pts	I
2	White-on-white	6pts	II
3	White-on-white	12pts	I
4	Black-on-white	8pts	II
5	White-on-white	8pts	I
6	Black-on-white	6pts	II
7	Black-on-white	8pts	I
8	White-on-white	8pts	II

ILLUSTRATION NO. 9
CAP PRESENTATION ORDER
1 TO 8 — PRESENTED TO CONSUMERS
1 TO 16 — PRESENTED TO THE EXPERTS



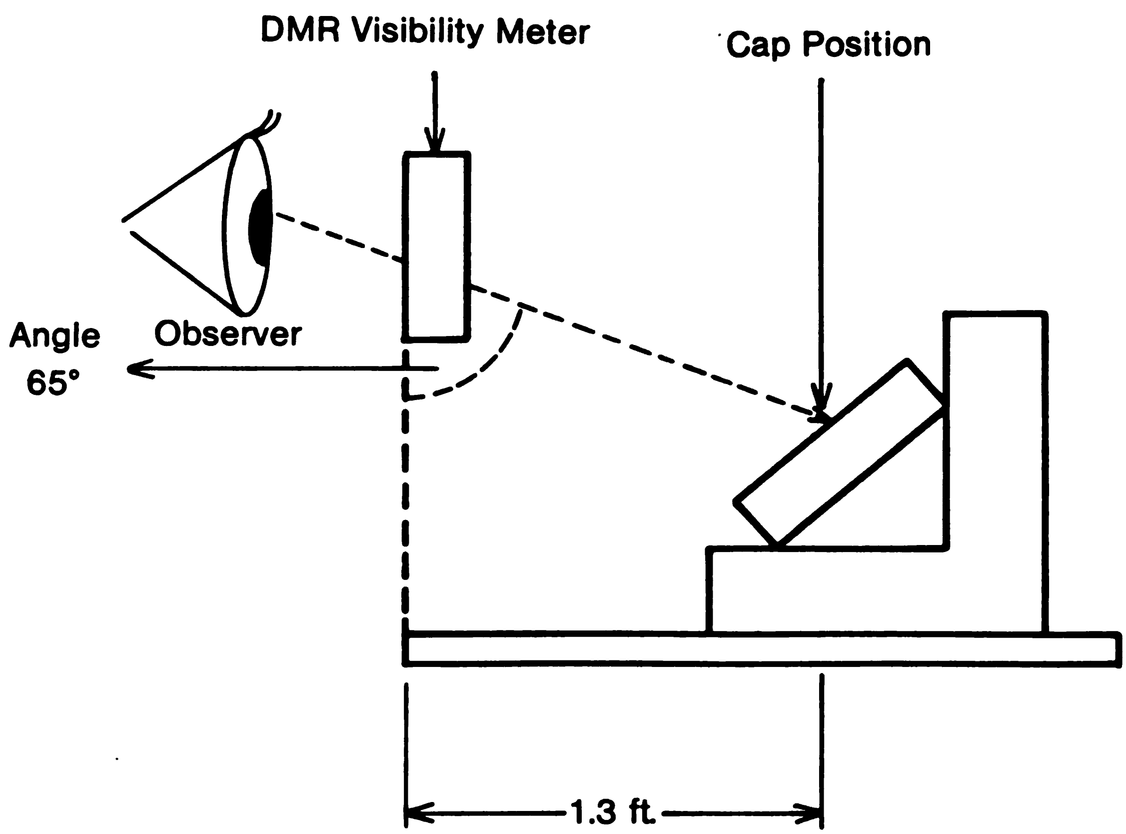
4.3. DMR Visibility Meter Procedure

The same subjects that were tested with the polariscope also took part in the visibility meter test. The test was individually administered under the same illumination of 50 foot candles as that of the polariscope. The visibility meter (see illustration No. 4 and 10) was held by a subject in front of the eyes, in approximately the same position that eye glasses are worn. On the table, 16 inches away from the subject's eyes a cap was placed on an L-shaped rigid polystyrene foam material such that it had a 45° inclination to help the subject read the message with ease.

At the beginning of the test, the density wedges were at the position of maximum opacity. In this case, the visual angle scale (in minutes) was at 20. Initially a trial test was conducted to familiarize the subject with the use of the meter and to build up confidence in its usage. The same sample cap was used with all the subjects. Once the sample cap was placed on the designated spot, the subject was instructed to rotate the thumb wheel on the visibility meter until he/she could see the message. There was no particular order of reading the message, but the subjects were encouraged to stop rotating the knob once they perceived the whole message.

As in the polariscopic procedure, the message on the cap was not revealed to the subject. Once the message was legible to the subject, the numerical reading (between 1 and 20) on the meter was recorded. After the familiarization

ILLUSTRATION NO. 10
A DIAGRAM SHOWING THE POSITION OF THE OBSERVER USING DMR VISIBILITY METER IN RELATION TO THE TEST CAP.



with the sample cap, it was replaced by the test cap and the same order of presentation (of cap number one to number eight) was followed (see Table 3). The visual acuity of the subjects was only recorded once, during the polariscopic test. Both tests were conducted consecutively with a break of about ten minutes to reduce the fatigue that might have been associated with the first test and to further diminish the memory of what the subject had seen in the previous test. The results were recorded on a test form (see Sample Test Form 1, page 51).

4.4. Expert Panel/Judge Procedure

An expert panel consisting of ten professionals was created as a standard whose evaluation could be compared to that of the consumers using the two devices. These experts were drawn from the professional field of graphics trades. They had a wide range of experience in this field, ranging from thirteen to thirty years, and their ages ranged from thirty two to sixty years. All of them had attended college.

The expert panel was asked to observe the caps and to give a numerical rating for legibility. Unlike the consumers, they did not use the two devices, but rather each cap was rated for legibility on a scale of 1 to 10. Ten was the highest level of legibility while one was the least. It was expected that this method of rating the caps would give some objective results which could be used in correlating the consumers data from the two devices with that of the expert judges as a reference standard. Initially before the test every expert was requested to fill out a questionnaire which included education/academic level, years of experience in a relevant field and title (see Sample Questionnaire No. 1, page 53).

The test was conducted in an ordinary office in the School of Packaging where the curtains were open and the level of illumination maintained at 50 foot candles. This setting was important because it represented a typical reading environment such as a home or office. Some of the office lighting was deliberately turned off to maintain the desired level of illumination.

At the beginning of the test each expert was familiarized with the subject matter by taking him/her for a visit to the test room consumers would use and informing him/her why the School of Packaging was conducting this study. Examples of child resistant bottle caps were shown in the laboratory with their varied messages and the objectives of the study on legibility were clearly defined. Each expert was also shown the two devices that the consumers would use in evaluating the legibility and how they would use them. In doing this, the importance of their contribution to this study was emphasized and the application of the results was explained.

In the testing room each expert was informed that sixteen caps would be rated for legibility, on a scale of 1 to 10, one at a time. Each cap was evaluated twice for legibility but the experts did not know this. They thought they were evaluating sixteen different caps. This is unlike the consumers who evaluated each cap once while using each of the two devices. The caps were presented to the experts in a fixed order (see illustration No. 9 and Table 4). The procedure was to have the expert seated at the table while each test cap from one to sixteen was presented one by one by the test monitor. There was no limit of time in which the experts could view any cap. The only emphasis was to place the cap 14-16 inches away from the eyes as this is usually the normal reading distance. The test was individually administered and the content of the message was not revealed to the experts before the test. After the expert rated each cap, he/she recorded the rating on a test form (see Sample Test Form No. 2, page 52). Before the beginning of the test, each expert read a Near Point Visual Acuity Card and the visual acuity was recorded on the aforementioned form. After evaluating all the caps, each expert was shown all of them together and relevant comments noted.

Table 4

Cap Order of Presentation to the Expert Panel

Cap No.	Color	Size	Message Type
1	Black-on-white	12pts	I
2	White-on-white	6pts	II
3	White-on-white	12pts	I
4	Black-on-white	8pts	II
5	White-on-white	8pts	I
6	Black-on-white	6pts	II
7	Black-on-white	8pts	I
8	White-on-white	8pts	II
9	Black-on-white	12pts	I
10	White-on-white	6pts	II
11	White-on-white	12pts	I
12	Black-on-white	8pts	II
13	White-on-white	8pts	I
14	Black-on-white	6pts	II
15	Black-on-white	8pts	I
16	White-on-white	8pts	II

SAMPLE TEST FORM ONE

Michigan State University
School of Packaging

Measurement of Legibility of CRP Caps

Data Sheet (A)
Consumers

No: ______________________ Visual Acuity: _____________

Age Group: Check One: Sex: ___________

____ 18-23

____ 30-55

Polariscope		DMR Visibility Meter	
Cap No.	Readings: Degrees of Rotation	Cap No.	Readings: Visual Angle in Minutes
1		1	
2		2	
3		3	
4		4	
5		5	
6		6	
7		7	
8		8	

SAMPLE TEST FORM TWO

Michigan State University
School of Packaging

Measurement of Legibility of CRP Caps

Data Sheet (B)
Expert Panel

No: ______________________ Visual Acuity ______________

Cap No.	Legibility Rating (On a Scale of 1-10)
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	

General Comments:

SAMPLE QUESTIONNAIRE FORM ONE

Michigan State University
School of Packaging
Expert Panel

We would like some information about you.

1. Name:

2. Work Address: City/Zip:

   Phone: ( )

3. Sex:

4. Age:

5. Title:

6. Education/academic Level:

7. Years of experience in printing/graphics/typography:

8. Have you ever taken part in a research involving an expert judge/panel? If yes, what type?

9. Would you be willing to participate in another research panel at some later date?

CHAPTER 5

DISCUSSION OF RESULTS

5.1. Introduction

In analyzing the results, the following questions relating to the objectives of this study will be discussed:

(i) Do the instruments give the same results as the experts? In other words, what is the correlation between the experts' rating of legibility on the eight caps with that of the consumers using the two instruments?

(ii) How do the two instruments compare?

(iii) What is the effect of color contrast, type size and visual acuity on the legibility of the two messages for (a) the experts and (b) the consumers using the two instruments?

5.2. Correlation Between the Experts' Rating and the Consumers Readings Using the Two Instruments.

The raw data for the experts' ratings of the sixteen cap presentation and the average rating of the eight cap variations is recorded in Appendix A page 90-100, while that of the consumers using the two instruments is recorded in Appendix B page 106-108.

Reliability of Experts

Before the correlation between instruments and experts was obtained, the data for the experts was analyzed for an estimate of reliability (or the degree of internal consistency) among the ten experts and their overall reliability for the sixteen caps.

In evaluating the expert data for reliability the method used for estimation was internal consistency by Cronbach Coefficient Alpha (α) (27). SPSS-X RELEASE 2.2 Program was used for this analysis. The following formula was used in determining the statistic r_{xx} (which represents reliability):

$$r_{xx} \text{ or } \alpha = \frac{n}{n-1}\left[1 - \frac{\Sigma S_i^2}{S_x^2}\right]$$

where n = number of items in test (e.g. 16 caps)

S_i^2 = variance of a single item (e.g. one cap)

Σ = Summation sign indicating that S_i^2 is summed over all items

S_x^2 = variance of the total test.

The reliability coefficients were reported for Standardized Item Alpha, which controls for variance of the measurements and is taken as the true estimate of reliability. A reliability coefficient of "0" would indicate a lack of any internal consistency among the experts, while "1.00" would indicate complete agreement. The reliability coefficients between pairs of experts (1 to 10, 2 to 10 etc. for all pairs) and the overall reliability for the ten experts is reported in Appendix A page 101.

The reliability coefficients between pairs ranged from -1.64 to .90, and the overall reliability among the experts was .83.

In their discussion of Reliability and Test Use, Mehrens and Lehmann (27) claim that if a measure is to be used to help make predictions about individuals, then it should be more reliable (high alpha value) than if it is to be used to make predictions about groups of people. It is also the two authors' opinion that standardized tests used to assist in making decisions about "individuals" should have reliability coefficients of at least .85, while for "group decisions," a reliability coefficient of about .65 may suffice. Similarly, Miles Tinker (23) says that for group comparisons, such as ordinarily employed in legibility studies, a reliability should not be less than .50; preferably, it should be .60 or above. The author continues to say that if one is dealing with individual diagnosis, the reliability coefficient should be at least .80, and preferably .90 or above.

For the purpose of this study, the measures or rating obtained from the experts will be used to help make decisions about legibility by using the consumers with 20/20 and 20/30 visual acuities using the two instruments. Therefore the overall reliability coefficient of .83

obtained from the ten experts is highly acceptable by the above standard and indicates a very high agreement or consistency among the experts. This internal consistency estimate will then be of value for interpreting the legibility scores of the CRP caps obtained from the consumers using the two instruments, which means that the experts can be relied upon as a reference control or standard for this test.

For the internal consistency method of estimating reliability, sources of error which may affect the estimates obtained are: sampling error, random error within the test, test length, speed, group homogeneity, item difficulty and objectivity, according to Mehrens and Lehmann (27). They further report that, the more heterogeneous the group, the higher the reliability and the more subjectively a measure is scored, the lower the reliability of the measure.

The expert panel was assembled from practitioners in graphic arts in order to reduce as much as possible the effect of subjectivity. Also, the panel was instructed about the purpose of the evaluation and they were asked to exercise their professional judgement as a further effort to reduce the effect of subjectivity. In the reliability analysis, a correlation matrix was also obtained. This data reinforces the results of the reliability coefficients, but was not used for this study. It is recorded in Appendix A, page 102 (between pairs of experts and all pairs).

Correlation between the experts and the instruments

Correlation coefficients (which measure how close a relationship is to a straight-line fit) were obtained between the experts' ratings and consumers' readings using the polariscope and DMR visibility meter. The data used to evaluate the correlation are reported in Tables 5-10 in order rankings for the above mentioned categories. The caps were ranked within each group from the most legible to least legible and whenever there was a tie in two values, the standard deviation value was used as the tie breaker, with the smaller deviation assumed to represent greater legibility. These rankings were used for comparisons between the experts and the consumers using the two instruments and for comparisons between the two instruments in order to find out if there was a pattern in the ranking order within these groups.

The method used to compute the correlation was based upon the Product-Moment Correlation (19). MSTAT Version 3 Program was used for this analysis. The following formula was used to determine the statistic r (Correlation Coefficient):

$$r = \frac{N\Sigma XY - \Sigma X \Sigma Y}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2]\,[N\Sigma Y^2 - (\Sigma Y)^2]}}$$

where N = number of caps (which were eight)

X and Y = ratio measures taken on N.

An example of the mathematical computation of correlation is shown in Appendix B page 110.

Table 5
Legibility Rating by Expert Panel vs. Consumer 20/20 Readings on Polariscope in Order Rankings

	Expert Panel					Polariscope				
Message	Rank Order[1]	Cap No.	Rating[2]	Color	Size (Pts)	Rank Order	Cap No.	Reading[2]	Color	Size (Pts)
I	1.	5	4	W/W	8	1.	5	9	W/W	8
	2.	3	6	W/W	12	2.	3	9	W/W	12
	3.	7	7	B/W	8	3.	7	6	B/W	8
	4.	1	9	B/W	12	4.	1	5	B/W	12
II	1.	2	2	W/W	6	1.	2	14	W/W	6
	2.	8	4	W/W	8	2.	8	10	W/W	8
	3.	6	6	B/W	6	3.	6	8	B/W	6
	4.	4	8	B/W	8	4.	4	6	B/W	8

1. Ranking Order is from 1 = least legible to 4 = most legible.

2. From Appendix A, page 100, mean rating for expert panel.

3. From Appendix B, page 107, mean reading for subjects using polariscope.

Table 6
Legibility Rating by Expert Panel vs. Consumer 20/30
Readings on Polariscope

	Expert Panel					Polariscope				
Message	Rank Order[1]	Cap No.	Rating[2]	Color	Size (Pts)	Rank Order	Cap No.	Reading[2]	Color	Size (Pts)
I	1.	5	4	W/W	8	1.	5	17	W/W	8
	2.	3	6	W/W	12	2.	3	13	W/W	12
	3.	7	7	B/W	8	3.	7	11	B/W	8
	4.	1	9	B/W	12	4.	1	8	B/W	12
II	1.	2	2	W/W	6	1.	2	24	W/W	6
	2.	8	4	W/W	8	2.	6	15	B/W	6
	3.	6	6	B/W	6	3.	8	14	W/W	8
	4.	4	8	B/W	8	4.	4	11	B/W	8

1. Ranking Order is from 1 = least legible to 4 = most legible.

2. From Appendix A, page 100, mean rating for expert panel.

3. From Appendix B, page 107, mean reading for subjects using polariscope.

Table 7
Legibility Rating by Expert Panel vs. Consumer 20/20
Readings on DMR Visibility Meter

	Expert Panel					DMR Meter				
Message	Rank Order[1]	Cap No.	Rating[2]	Color	Size (Pts)	Rank Order	Cap No.	Reading[2]	Color	Size (Pts)
I	1.	5	4	W/W	8	1.	5	11	W/W	8
	2.	3	6	W/W	12	2.	3	14	W/W	12
	3.	7	7	B/W	8	3.	7	17	B/W	8
	4.	1	9	B/W	12	4.	1	19	B/W	12
II	1.	2	2	W/W	6	1.	2	10	W/W	6
	2.	8	4	W/W	8	2.	8	13	W/W	8
	3.	6	6	B/W	6	3.	6	14	B/W	6
	4.	4	8	B/W	8	4.	4	17	B/W	8

1. Ranking Order is from 1 = least legible to 4 = most legible.

2. From Appendix A, page 100, mean rating for expert panel.

3. From Appendix B, page 108, mean reading for subjects using DMR visibility meter.

Table 8
Legibility Rating by Expert Panel vs. Consumer 20/30
Readings on DMR Visibility Meter

	Expert Panel					DMR Meter				
Message	Rank Order[1]	Cap No.	Rating[2]	Color	Size (Pts)	Rank Order	Cap No.	Reading[2]	Color	Size (Pts)
	1.	5	4	W/W	8	1.	5	10	W/W	8
I	2.	3	6	W/W	12	2.	3	13	W/W	12
	3.	7	7	B/W	8	3.	7	15	B/W	8
	4.	1	9	B/W	12	4.	1	19	B/W	12
	1.	2	2	W/W	6	1.	2	9	W/W	6
II	2.	8	4	W/W	8	2.	8	11	W/W	8
	3.	6	6	B/W	6	3.	6	13	B/W	6
	4.	4	8	B/W	8	4.	4	16	B/W	8

1. Ranking Order is from 1 = least legible to 4 = most legible.

2. From Appendix A, page 100, mean rating for expert panel.

3. From Appendix B, page 108, mean reading for subjects using DMR visibility meter.

Table 9
Legibility Readings on Polariscope vs. Readings on DMR Visibility
Meter by Consumer 20/20 in Order Rankings

	Polariscope					DMR Meter				
Message	Rank Order[1]	Cap No.	Rating[2]	Color	Size (Pts)	Rank Order	Cap No.	Reading[2]	Color	Size (Pts)
I	1.	5	9	W/W	8	1.	5	11	W/W	8
	2.	3	9	W/W	12	2.	3	14	W/W	12
	3.	7	6	B/W	8	3.	7	17	B/W	8
	4.	1	5	B/W	12	4.	1	19	B/W	12
II	1.	2	14	W/W	6	1.	2	10	W/W	6
	2.	8	10	W/W	8	2.	8	13	W/W	8
	3.	6	8	B/W	6	3.	6	14	B/W	6
	4.	4	6	B/W	8	4.	4	17	B/W	8

1. Ranking Order is from 1 = least legible to 4 = most legible.

2. From Appendix B, page 107, mean reading for subjects using polariscope.

3. From Appendix B, page 108, mean reading for subjects using DMR visibility meter.

Table 10
Legibility Readings on Polariscope vs. Readings on DMR Visibility Meter by Consumer 20/30 in Order Rankings

	Polariscope					DMR Meter				
Message	Rank Order[1]	Cap No.	Rating[2]	Color	Size (Pts)	Rank Order	Cap No.	Reading[2]	Color	Size (Pts)
I	1.	5	17	W/W	8	1.	5	10	W/W	8
	2.	3	13	W/W	12	2.	3	13	W/W	12
	3.	7	11	B/W	8	3.	7	15	B/W	8
	4.	1	8	B/W	12	4.	1	19	B/W	12
II	1.	2	24	W/W	6	1.	2	9	W/W	6
	2.	6	15	B/W	6	2.	8	11	W/W	8
	3.	8	14	W/W	8	3.	6	13	B/W	6
	4.	4	11	B/W	8	4.	4	16	B/W	8

1. Ranking Order is from 1 = least legible to 4 = most legible.

2. From Appendix B, page 107, mean reading for subjects using polariscope.

3. From Appendix B, page 108, mean reading for subjects using DMR visibility meter.

A correlation coefficient of 1.00 indicates a perfect positive relationship or agreement while that of -1.00 indicates a perfect negative relationship or complete disagreement. In the discussion on correlation Theodore Horvath (19) claims that before the result can be interpreted, the r coefficient must be tested for significance. The test of significance establishes whether, at some confidence level, the present outcome should be viewed as the result of a sampling error. A test of significance for the r coefficient was computed before the interpretation of the results was made. The r coefficient was tested by comparison with the r critical value at the 0.05 significance level obtained from Table H, page 360, of "Basic Statistics for Behavioral Sciences" (Theodore Horvath) (20).

The following correlations were tested for r critical:

(i) Experts vs. Consumers 20/20 using polariscope

(ii) Experts vs. Consumers 20/30 using polariscope

(iii) Experts vs. Consumers 20/20 using DMR meter

(iv) Experts vs. Consumers 20/30 using DMR meter

(v) Polariscope vs. DMR meter using Consumers 20/20

(iv) Polariscope vs. DMR meter using Consumers 20/30

All the null hypotheses are stated as: no difference/no effect. The null hypotheses are rejected when the calculated r equals or exceeds the critical value obtained from Table H. These hypotheses are tested at the 0.05 significance level. The value of the statistic r, r critical and the decision on statistically significant difference are recorded in Table 11.

Reviewing the results of this study from Table 11, item (1), Experts vs. polariscope 20/20, shows a strong negative relationship between the expert ratings and the readings for consumers with 20/20 visual acuity on the polariscope. This relationship is such that the higher ratings of legibility by the experts is associated with lower readings on the polariscope by the consumers with 20/20 visual acuity. The same is true of number (2), Experts vs. polariscope 20/30. Figures 1 and 2 show bivariate plots for the data given in Tables 5 and 6. This strong correlation is further corroborated by Tables 5 and 6 where the caps' order ranking by consumers with 20/20 visual acuity is the same as that of the experts. The order ranking changes slightly for the 20/30 consumer group when reading message II (Table 6). These subjects saw the 8 point, white-on-white print as slightly easier to read than the black on white 6 point print.

Table 11, item 3 and 4, experts vs. DMR visibility meter 20/20 and 20/30 shows a strong positive correlation between the experts' rating of legibility and the DMR visibility meter readings when viewed by the consumers with 20/20 and 20/30 visual acuity. The relationship is such that the higher ratings of legibility by the experts are associated with higher readings on the DMR visibility meter by the consumers. This relationship is corroborated by

Table 11
Correlation Values and Statistical Significance for Expert Panel Ratings and Consumer Panel Machine Readings (Horvath Test of Significance with r Critical = 0.707).

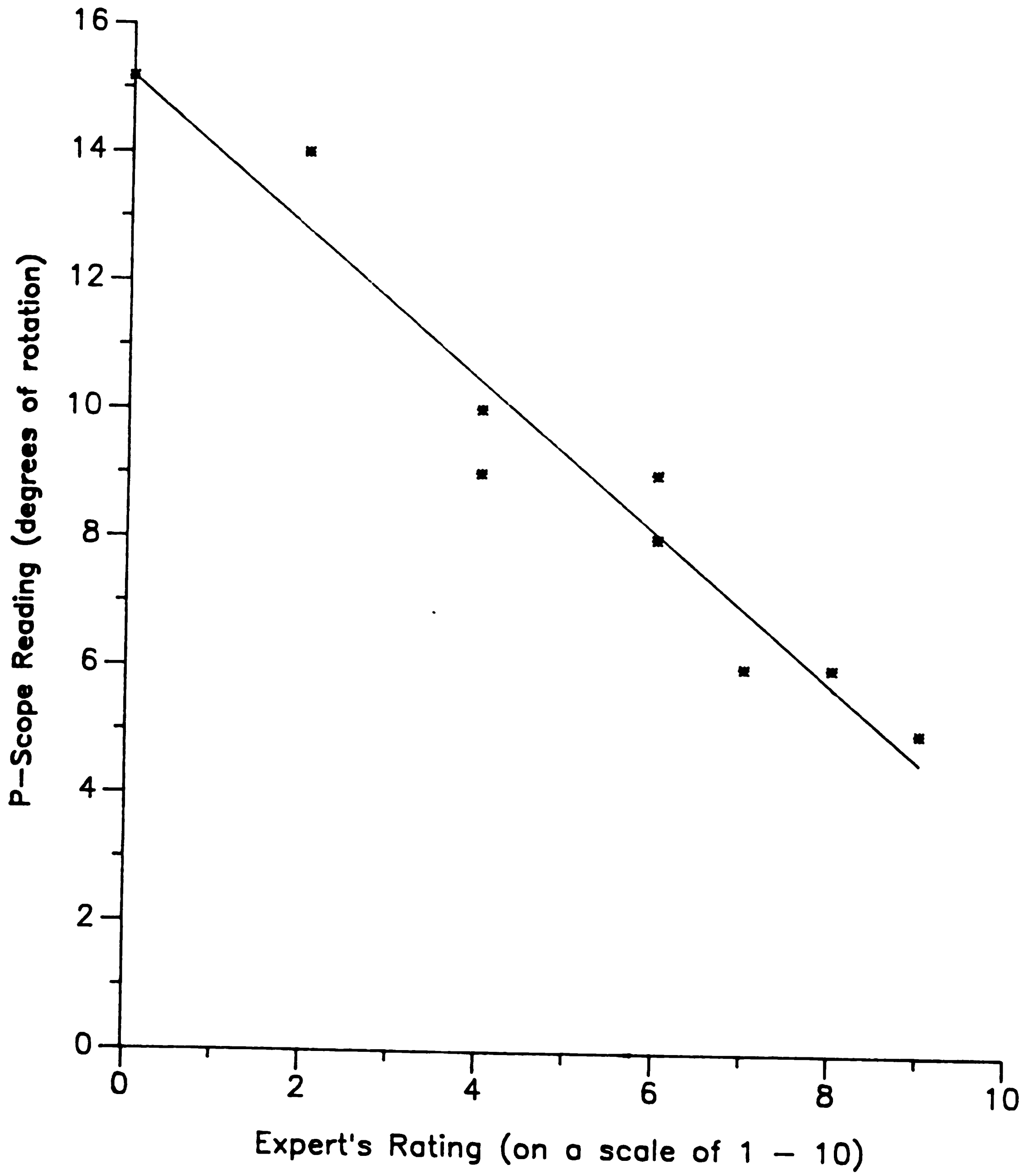
Source of Correlation	r Coefficient	Statistical Significance at 0.05 significance level
1. Experts vs. M_1 20/20	-0.950	Yes
2. Experts vs. M_1 20/30	-0.951	Yes
3. Experts vs. M_2 20/20	0.966	Yes
4. Experts vs. M_2 20/30	0.974	Yes
5. M_1 vs. M_2 20/20	-0.911	Yes
6. M_1 vs. M_2 20/30	-0.905	Yes

df = N-2 = 8-2 = 6 (8 caps were observed).

M_1 = polariscope viewed by Consumers with the given visual acuity.

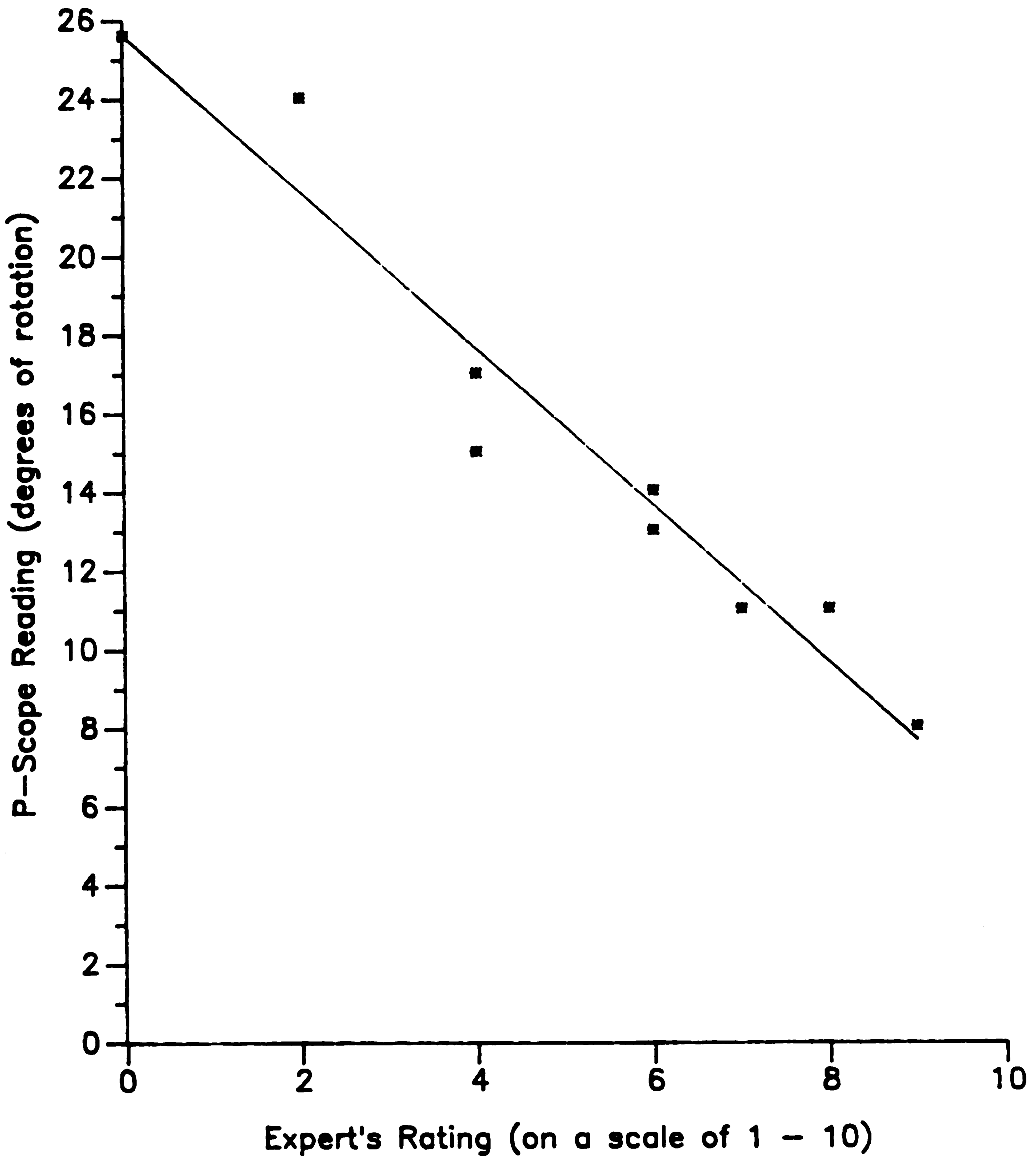
M_2 = DMR visibility meter viewed by Consumers with the given visual acuity.

Figure 1
Bivariate Plot showing correlation between the expert's rating and polariscope reading by consumers with 20/20 visual acuity.



Correlation = −0.950
Intercept = 15.160
Slope = −1.180

Figure 2
Bivariate Plot showing correlation between the expert's rating and polariscope reading by consumers with 20/30 visual acuity.



Correlation = −0.951
Intercept = 25.590
Slope = −1.993

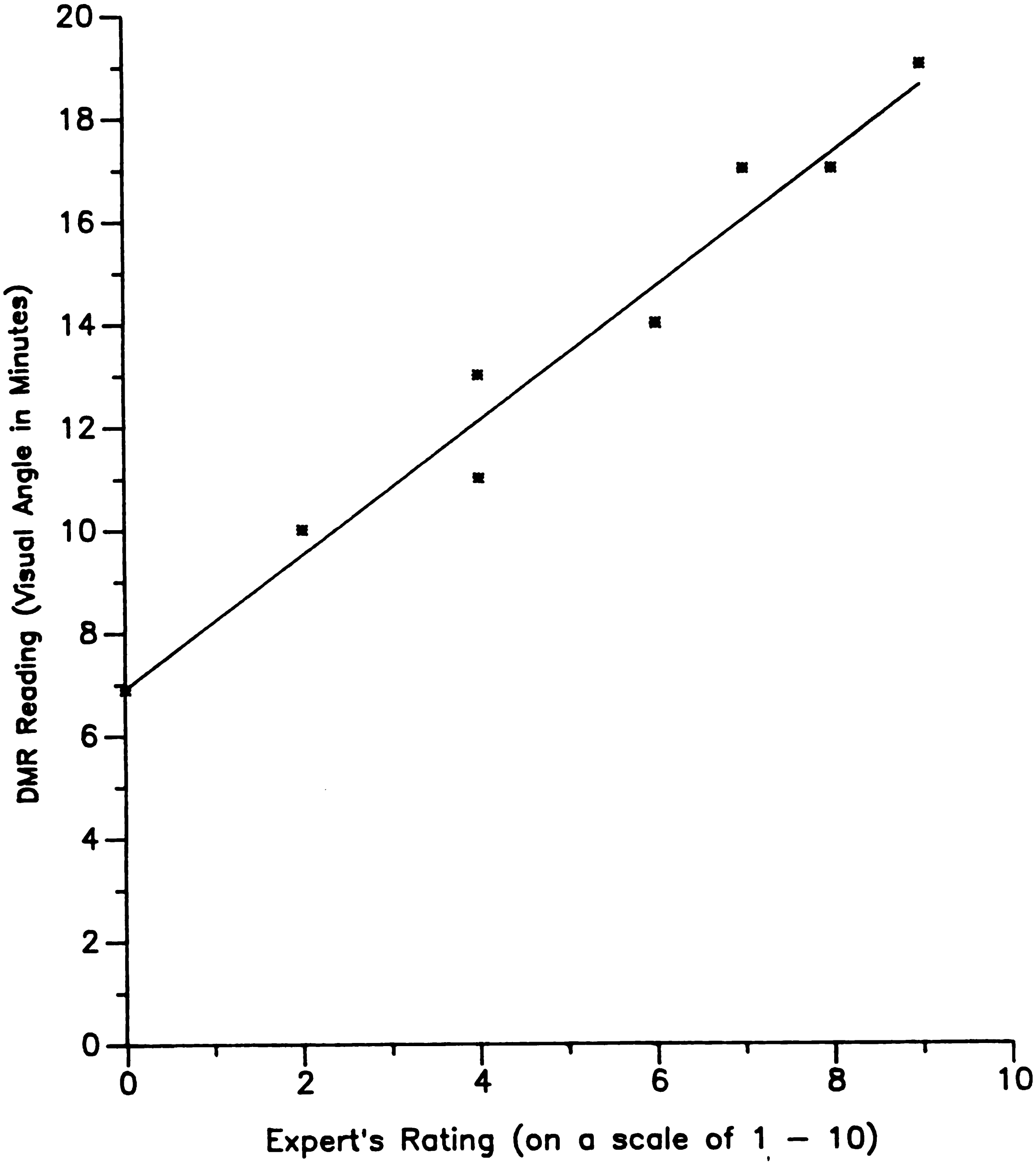
Tables 5 and 6 where the caps' order rankings for consumers with the two visual acuities are in total agreement with that of the experts. Figures 3 and 4 show bivariate plots for the data given in Tables 7 and 8 illustrating these correlations. It is evident from the bivariate plots that a stronger relationship exists between the experts and DMR visibility meter 20/30 than 20/20, as shown by the closeness of values to the regression line.

From the analysis of Table 11 and the subsequent discussions on correlation it can be stated that the DMR visibility meter readings agree well with the ratings given by the experts. This is reinforced by the order rankings for the caps in Tables 7 and 8 in which consumer rankings follow exactly those of the experts. Since the experts' reliability was high, the high degree of correlation of experts with consumers was taken as at least partial validation for examining the correlation between the instruments.

5.3. Comparison (Correlation) of the Polariscope and DMR Visibility Meter

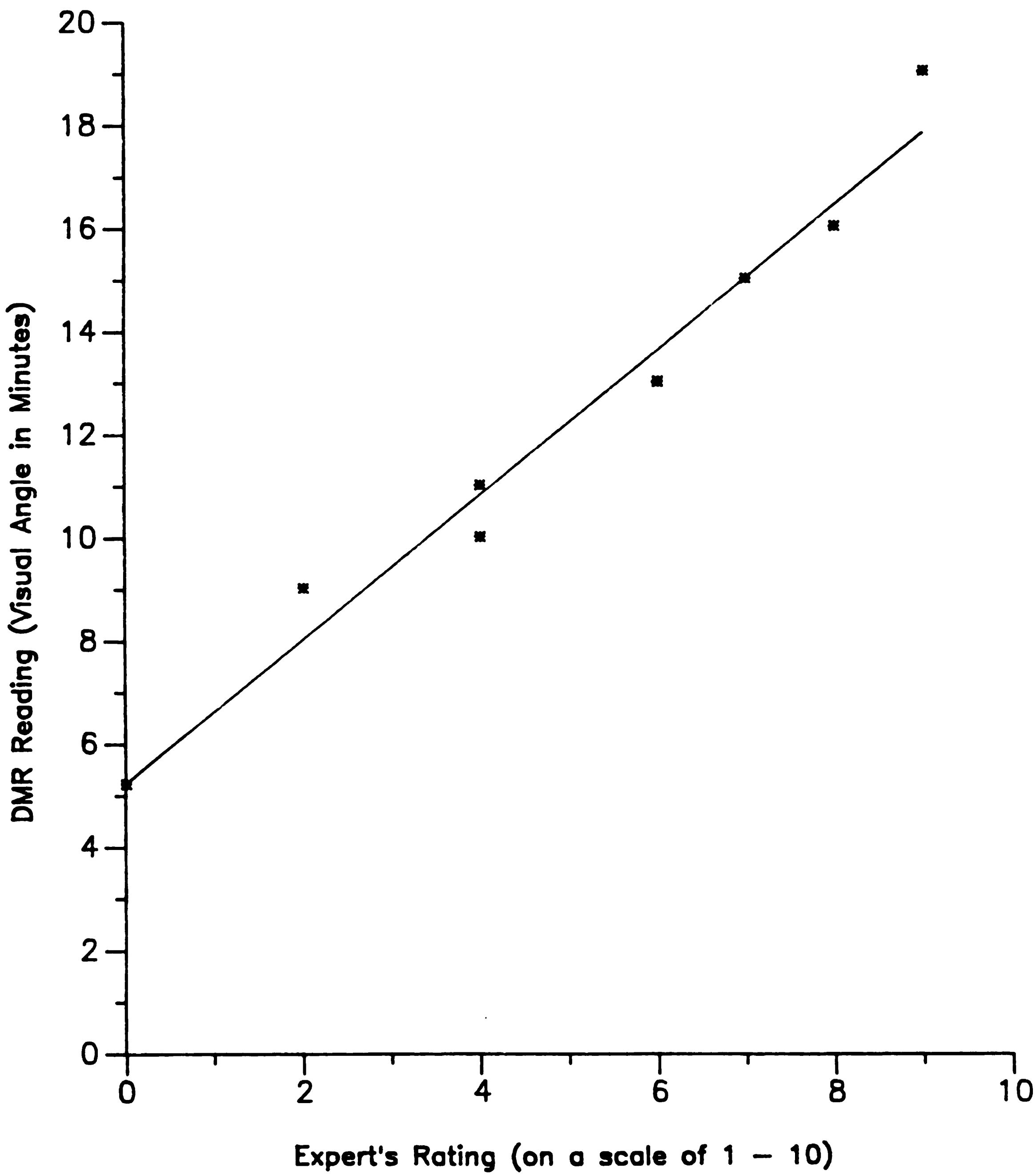
From the correlations reported in Table 11, there is a very strong negative relationship between the polariscope and DMR visibility meter at both visual acuities. This relationship is such that the higher readings of legibility by consumers on polariscope is associated with lower readings on the DMR visibility meter. The negative relationship is explained by the difference in scales of

Figure 3
Bivariate Plot showing correlation between the expert's rating and DMR visibility meter reading by consumers with 20/20 visual acuity.



Correlation = 0.966
Intercept = 6.900
Slope = 1.300

Figure 4
Bivariate Plot showing correlation between the expert's rating and DMR visibility meter reading by consumers with 20/30 visual acuity.



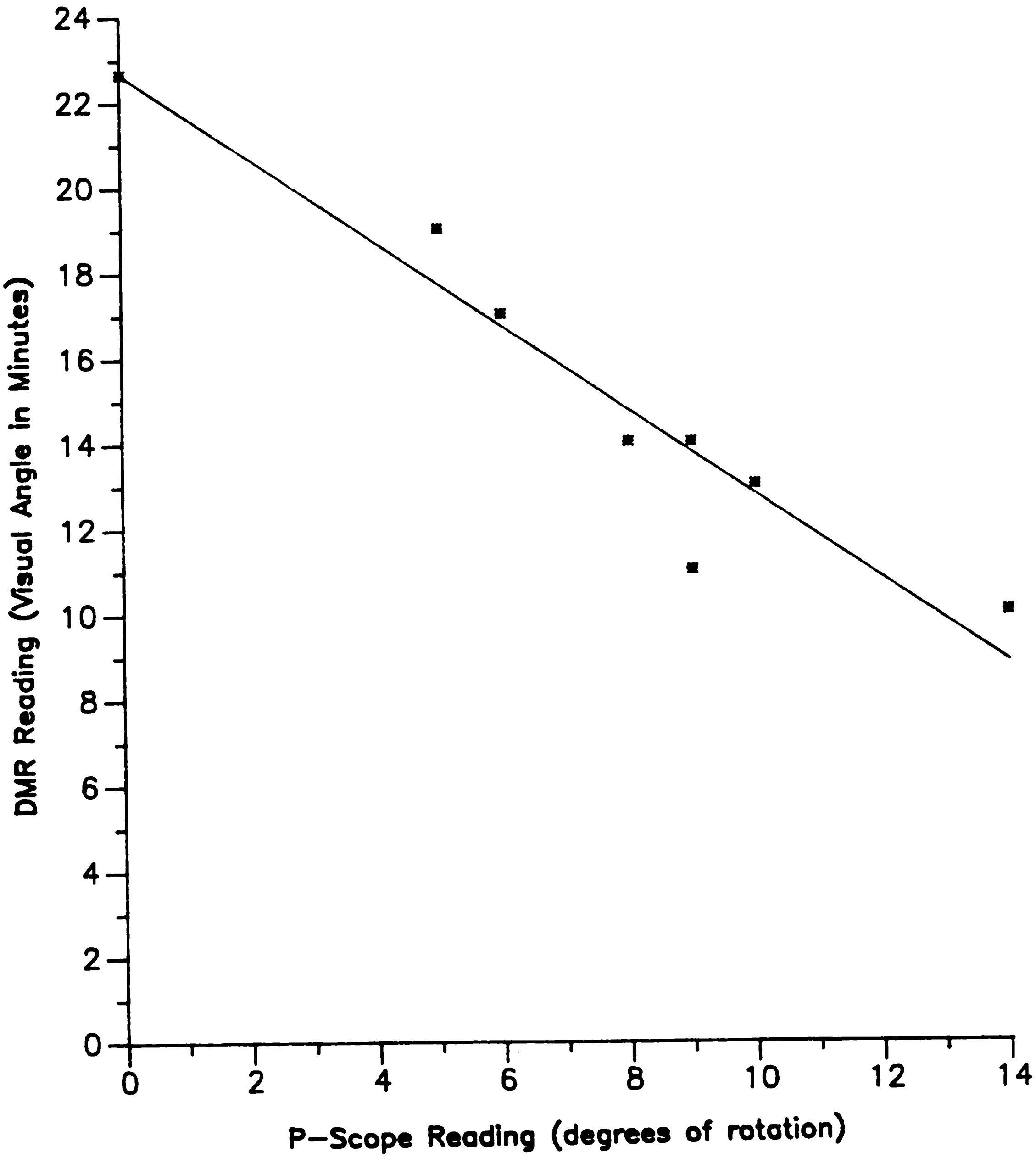
Correlation = 0.974
Intercept = 5.200
Slope = 1.400

these instruments. The scales are reversed, with a high legibility value for polariscope is associated with a low number (small rotation) while on the DMR visibility meter, it is associated with a high number (large rotation of the density wedge). The polariscope's scale ranges from 0-90 degrees while that of the DMR visibility meter is 1-20 and therefore is much more compressed than the polariscope scale.

Tables 9 and 10 show the order ranking for legibility of the eight caps for both instruments, with the consumers group 20/20 and 20/30. The cap order ranking for both instruments is the same for the consumers 20/20 group and this is reinforced by the high correlation (-0.911) while that of the 20/30 is slightly lower (-0.905) since the order ranking changes slightly. Once again, it is for message II, with this group that the ranking is different for the polariscope than for either experts or the DMR visibility meter.

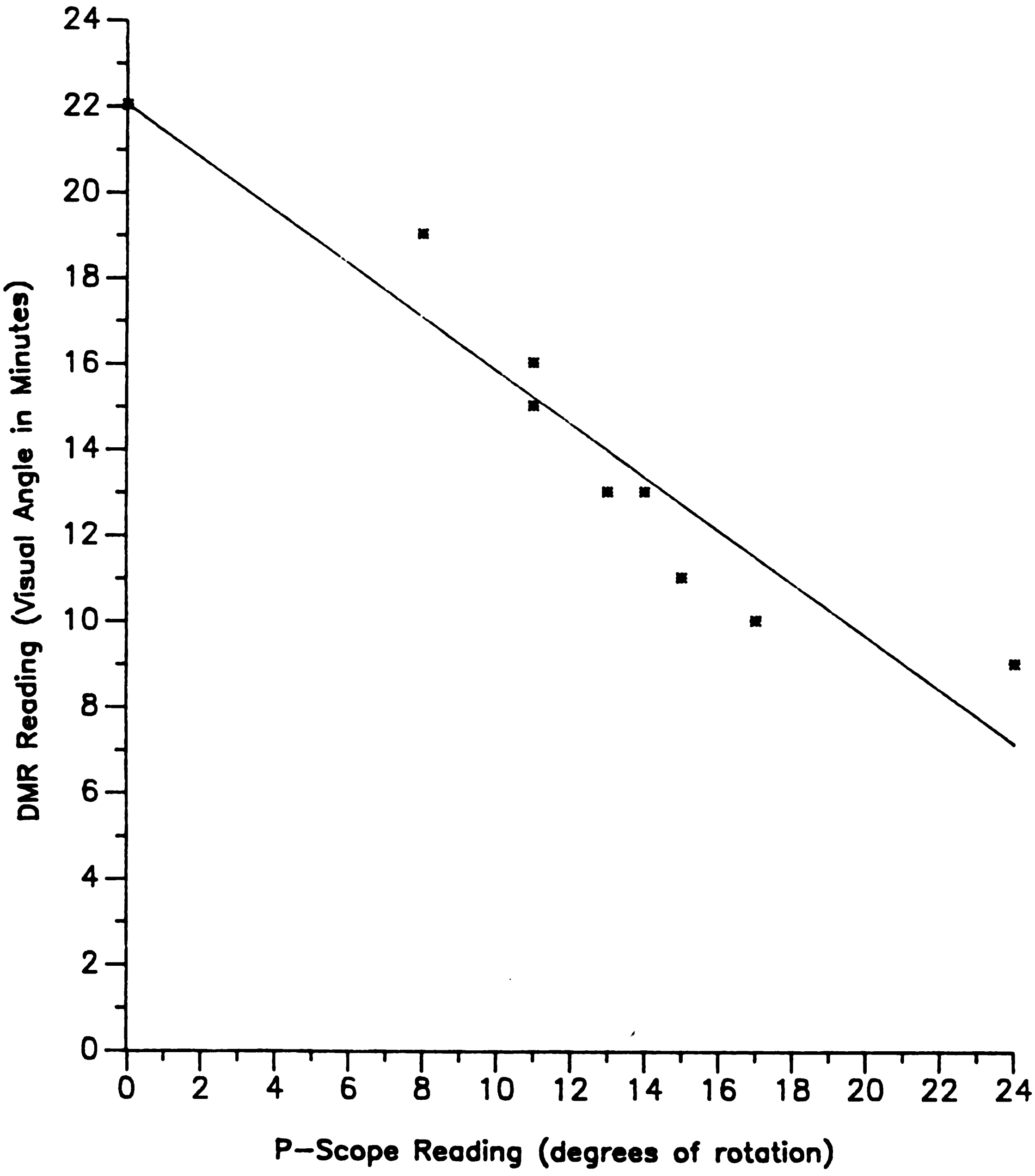
Figures 5 and 6 show the negative linear relationship between the polariscope and DMR visibility meter for the data from Table 9 and 10, for consumers with the two visual acuities. From the correlations obtained between the experts and both of the instruments, it is evident that both can be used to evaluate legibility since their scale readings agree with the ratings obtained from the experts with the same relative numerical order of the caps despite the reversed scales. The question that arises then, is how

Figure 5
Bivariate Plot showing correlation between the Polariscope reading and DMR visibility meter reading by consumers with 20/20 visual acuity.



Correlation = −0.911
Intercept = 22.640
Slope = −0.987

Figure 6
Bivariate Plot showing correlation between the Polariscope reading and DMR visibility meter reading by consumers with 20/30 visual acuity.



Correlation = −0.905
Intercept = 22.010
Slope = −0.620

to rate their performance. Performance of the two instruments can be evaluated and rated only if their individual reliabilities are established, according to Mr. Joshua G. Bagaka (2), Statistical Consultant, Research and Evaluation Services, at Michigan State University. After computing their reliabilities or consistency, the two instruments can then be compared and rated. Reliability analysis requires at least two measures of the same thing from a random population. This was done for the experts, but the legibility test of the eight CRP caps was performed only once for each instrument and the data collected was not adequate to evaluate the instruments' reliability. It is therefore not feasible to compare the two instruments' performance with the available data and this should be a subject for another investigation.

The polariscope has the advantage of an extended scale (1-90 degrees) unlike the DMR visibility meter's which is compressed (1-20 visual angle in minutes). Due to the compressed scale, more light tends to be allowed in per increment of rotation, while viewing a target. In polariscope the light transmission increase tends to be more gradual which results in a relatively long scale and small change in light per increment (greater sensitivity). Consumers interviewed immediately after using the two instruments tended to be more comfortable with the polariscope than with the DMR visibility meter. They felt it was easier to manipulate and the wide viewing screen did not strain the users as did the DMR visibility meter with its binocular structure. One advantage of the DMR visibility meter over the polariscope is that it can be administered to consumers in a wider variety of situations because it is lighter in weight and less bulky. The polariscope is designed to be used primarily in the laboratory where illumination can easily be controlled.

5.4. Effect of Message, Color Contrast, Type Size and by Visual Acuity on the Legibility of the CRP Caps for the Experts and Consumers (Using the Two Instruments)

The statistical evaluation of the above effects was based on a Split-Plot Factorial design, Type SPF-pru.q design (6) utilizing analysis of variance (ANOVA). SAS.A1 Program was used for this analysis. There were four treatments or main effects and eleven treatment combinations or interactions of the main effects for the consumers using the two instruments. In the case of the experts, there were three main treatments and four treatment combinations for the examination of interaction effects. Table 12 shows the main treatments and the treatment combination for the experts and consumers in this test. Visual acuity for the experts was not important in this test. The layout or block diagram of this design (SPF - 222.4 Design) and its structural model (6), is shown in Appendix B, page 109.

Table 12
Main Treatments and Treatment Combinations

Source	Treatments		Possible Combinations
Consumers	Message	(A)	AB, AC, AD, ABC, ABD ACD and ABCD
	Visual	(B)	BC, BD, and BCD
	Color	(C)	CD
	Size	(D)	-
Experts	Message	(A)	AB, AC, and ABC
	Color	(B)	BC
	Size	(C)	-

The ANOVA computational procedure (6) was used to evaluate the significance of the main and treatment combinations in Table 12. The ANOVA table for the Experts is shown in Appendix A, page 103, while that of the consumers using the two instruments is shown in Appendix B, page 111, for the polariscope and page 115 for the DMR visibility meter. In this test, all the main effects were significant (except for visual acuity on the DMR visibility meter) and are reported in Table 13 in summary form.

In calculating the F value, all the null hypotheses were stated as: no difference/no effect.

The null hypotheses are rejected when the probality > F value is equal to or less than 0.05. Table 13 shows that all the main effects were significant for both the experts

Table 13
ANOVA Table for the Significant Main Effects

Test Group	Source of Variation	DF	ANOVA SS	F value	Statistical Significance at 0.05 significance level
Experts[1]	Message	1	45.00	22.59	Yes
	Color	1	248.51	124.77	Yes
	Size	2	129.06	32.40	Yes
Consumers[2] (polariscope)	Message	1	333.51	11.62	Yes
	Visual	1	971.80	33.85	Yes
	Color	1	888.31	30.94	Yes
	Size	2	810.51	14.11	Yes
Consumers[3] (DMR Visibility Meter)	Message	1	126.03	10.77	Yes
	Visual	1	43.89	3.75	No
	Color	1	960.40	82.08	Yes
	Size	2	432.85	18.50	Yes

1. All main effects significant at 0.9999 confidence level.
2. Message significant at 0.9992 confidence level. All the other effects significant at 0.9999 confidence level.
3. Message significant at 0.9987 confidence level. For visual acuity, α = .0547, so this effect should be noted along with the main effects. All the other effects significant at 0.9999.

and the consumers apart from visual acuity on the DMR visibility meter. This means that message, color contrast, size and visual acuity differences affected the legibility, while the interactions did not significantly affect the legibility.

When ANOVA shows significance between two treatments it is not necessary to do further testing as interpretations can be made directly, but when the ANOVA shows significance for three or more treatments, it is necessary to do a post-F comparison procedure in order to identify which treatments differ significantly. There are many such tests and Turkey's HSD (Honestly Significant Difference) test (21) is one of them. According to Theodore Horvath (21), Tukey's HSD procedure would be applied by determining the critical difference between independent variable level totals or pairwise comparisons among means. SAS.A1 Program was used for this analysis. A comparison involving two means is declared to be significant if it exceeds HSD, which is given by :

$$HSD = q_{\alpha, v} \sqrt{\frac{MS_{error}}{n}}$$

where q is obtained from the distribution of the studentized range statistic as given in Table D.7.1 of Statistical methods in Education and Psychology, Glass/Hopkins. (7)

α = significance level, in this case, 0.05

v = degree of Freedom

MSE = Mean Square Error Term

n = number of scores in each group

An example of the mathematical computation of HSD is shown in Appendix A page 105.

Even though the ANOVA values for message, color and visual acuity were significant (see Table 13), and allowed direct interpretation immediately, Tukey HSD analysis results are also reported for them (Tables 14, 15 and 16). These results support the ANOVA values, but are infact redundant and not required. However, there were three treatments for type size which were significant and for these ANOVA alone did not identify the individual treatments that were significantly different. It was therefore necessary to do a Tukey HSD test to determine which levels between the three type sizes differed significantly. Turkey's Studentized Range (HSD) test for experts and consumers is reported in Tables 14, 15, and 16 in summary form. The legibility ratios of the means for the pairs of variables are reported in Tables 17 and 18. The full analysis is shown in Appendix A, page 103 to 105 for the experts, and Appendix B, page 112 to 114 for consumers using polariscope and page 116 to 118 for consumers using the DMR visibility meter. The null hypothesis adopted for this test was: no difference/no effect. The hypotheses were tested at the 0.05 significance level. Any hypothesis was rejected if the difference between the means exceeded the critical value

Table 14
Results of Tukey's HSD Test for Comparison of the Means of the Main Effects for the Experts

Source of Comparison	Means	Difference Between Means	HSD* Statistic	Statistical Significance at .05 Level
Message 1	6.3375	1.500	0.6290	Yes
Message 2	4.8375			
Color 1 - B/W	7.3500	3.5250	0.6290	Yes
Color 2 - W/W	3.8250			
Size 1 - 6pts	-	1.5375	1.5102	Yes
Size 2 - 8pts	-			
Size 1 - 6pts	-	3.5750	1.5102	Yes
Size 3 - 12pts	-			
Size 2 - 8pts	-	2.0375	1.5102	Yes
Size 3 - 12pts	-			

* The HSD statistic for message and color reported from the "Minimum Significance Difference" values shown in Appendix A page 104. The HSD statistic for the type sizes is calculated using the HSD formula by, using the values shown in Appendix A, page 105.

Table 15
Results of Tukey's HSD Test for Comparison of the Means of the Main Effects for the Consumers Using Polariscope

Source of Comparison	Means	Difference Between Means	HSD* Statistic	Statistical Significance at .05 Level
Message 1	8.4375	2.8875	1.6746	Yes
Message 2	11.3250			
Color 1 - B/W	7.5250	4.7120	1.6746	Yes
Color 2 - W/W	12.2375			
Visual 1 - 20/20	8.2679	5.3779	1.8271	Yes
Visual 2 - 20/30	13.6458			
Size 1 - 6pts	-	6.7948	5.6746	Yes
Size 2 - 8pts	-			
Size 1 - 6pts	-	6.1000	5.6746	Yes
Size 3 - 12pts	-			
Size 2 - 8pts	-	1.7625	5.6746	No
Size 3 - 12pts	-			

* The HSD statistic for message, color and visual is reported from the "Minimum Significance Difference" values, shown in Appendix B, page 112 to 113. The HSD statistic for the type sizes is calculated using the HSD formula by using the values shown in Appendix B, page 114.

Table 16
Results of Tukey's HSD Test for Comparison of the Means of the Main Effects for the Consumers Using DMR Visibility Meter

Source of Comparison	Means	Difference Between Means	HSD* Statistic	Statistical Significance at .05 Level
Message 1	14.5625	1.7750	1.0690	Yes
Message 2	12.7875			
Color 1 - B/W	16.1250	4.9000	1.0690	Yes
Color 2 - W/W	11.2250			
Visual 1 - 20/20	14.0179	1.1429	1.1664	No
Visual 2 - 20/30	12.8750			
Size 1 - 6pts	-	3.9937	3.6227	Yes
Size 2 - 8pts	-			
Size 1 - 6pts	-	4.6500	3.6227	Yes
Size 3 - 12pts	-			
Size 2 - 8pts	-	3.7937	3.6227	Yes
Size 3 - 12pts	-			

* The HSD statistic for message, color and visual is reported from the "Minimum Significance Difference" values shown in Appendix B, page 116 to 117. The HSD statistic for the type sizes is calculated using the HSD formula by using the values shown in Appendix B, page 118.

of the studentized range or HSD. Therefore the comparison involving the two means was declared significant.

An examination of Tables 14, 15, and 16, indicates that the three pairwise comparisons of the two messages are statistically significantly different. Tables 17 and 18 show the ratios of legibility which demonstrated the practical importance of the differences. Message I was almost 1.3 times more legible to the experts than message II (see Table 17). This difference is logical when one considers that the message content and message layouts were different on the two sets of caps evaluated. The difference is greater with the polariscope than the DMR visibility meter (1.3 and 1.1 times respectively). The data in Appendix A, page 100 on experts' average ratings reveals that, as the message II became harder to read, the standard deviation increased.

Color contrast greatly affected the legibility of CRP caps as the results of this test showed (see Table 17). For the experts, it was twice as easy to read the caps printed in black-on-white as the white-on-white. This is also supported by the results in Table 14 where the difference between the two colors was greatly significant. This was also true for the consumers using the two instruments (Tables 15, 16 and 17). It was 1.6 and 1.5 times easier to read the print in black-on-white than white-on-white with the polariscope and DMR visibility meter respectively, and

Table 17
Legibility Ratios* of the Means for the Pairs of Variables

	Legibility		
Variables	Experts Ratio	Polariscope Ratio	DMR Visibility Meter Ratio
Message	1.3X	1.3X	1.1X
Color Contrast	2.0X	1.6X	1.5X
Visual Acuity	-	1.7X	1.1X

For message, the ratio is:

$$\frac{\text{legibility of message I}}{\text{legibility of message II}}$$

For color contrast, the ratio is:

$$\frac{\text{legibility of Black-on-White}}{\text{legibility of White-on-White}}$$

For visual acuity, the ratio is:

$$\frac{\text{legibility for 20/20}}{\text{legibility for 20/30}}$$

Table 18
Type Size Legibility Ratios* for Message I and II Means

	Message I				Message II			
	Color	Size	Mean	Ratio	Color	Size	Mean	Ratio
Experts	W/W	8pts	4	1.5X	W/W	6pts	2	2.0X
	W/W	12pts	6		W/W	8pts	4	
	B/W	8pts	7	1.3X	B/W	6pts	6	1.3X
	B/W	12pts	9		B/W	8pts	8	
Polariscope	W/W	8pts	6	0.67X	W/W	6pts	14	1.4X
	W/W	12pts	9		W/W	8pts	10	
	B/W	8pts	6	1.1X	B/W	6pts	8	1.3X
	B/W	12pts	5		B/W	8pts	6	
DMR visibility meter	W/W	8pts	17	0.82X	W/W	6pts	10	1.3X
	W/W	12pts	14		W/W	8pts	13	
	B/W	8pts	17	1.1X	B/W	6pts	14	1.2X
	B/W	12pts	19		B/W	8pts	17	

* For type size, the ratio is:

$$\frac{\text{legibility of large type size}}{\text{legibility of smaller type size}}$$

the difference was greater in the polariscope than the DMR visibility meter.

Legibility was affected by visual acuity of the consumers tested using the polariscope but not the DMR visibility meter. In general, Tables 9 and 10 for the polariscope and DMR visibility meter order rankings show that the consumers with 20/20 visual acuity found it easier to read the two messages than those with 20/30. Table 15, for polariscope, shows a significant difference of the two acuities. According to the legibility ratios in Table 17, the consumer group with 20/20 visual acuity found it easier to read the messages on the CRP caps by more than one and a half times. For the DMR visibility meter the difference between the two acuities was almost significant at the 0.05 level (See the ANOVA Table in Appendix B, page 115). This is reinforced by the ratio on Table 17, which shows the consumer group with 20/20 visual acuity able to read the cap messages 1.1 times easier than those with 20/30. The reason for this small difference between the two acuities can probably be accounted for by the compressed scale of this instrument. The readings tended to be very close together and this was reflected by the very small difference between the means of the two visual acuities.

The effect of type size on legibility was evaluated. A comparison was made between 6 and 8 pts, 6 and 12 pts and 8 and 12 pts (Tables 14, 15, 16 and 18) to see whether significant differences existed. There is a significant difference between 6 pts and 8 pts for the experts and the consumers with the two instruments. Table 18 shows a high ratio of legibility for message II which ranged from 1.2 to 2 times for the experts, polariscope and DMR visibility meter. The 8 pt size was easier to read than the 6 pt size under the same conditions. The greatest difference is shown between 6 pts and 12 pts for the experts and the consumers with the DMR visibility meter (See Tables 14 and 16). This may not mean alot since we are comparing 12 pts in message I and 6 pts in message II whose type faces, message content and layout were different, but in general these differences have been demonstrated, especially where uniform type sizes have been used (8). The difference between 8 pts and 12 pts is significant for the experts and consumers using the DMR visibility meter, but not the polariscope. In the two cases above, the 12 pt message was easier to read than the 8 pt message, but in the case for polariscope, the two type size difference did not significantly affect the legibility results. This can be attributed to the closeness of values obtained for both 12 and 8 pt sizes. This is further supported by the order rankings on Table 9 where the ratings or scores for 8 and 12 pts were very close. The size anomaly between 8 and 12 pts is shown on Table 18, where the ratio for the white-on-white variation for both instruments is less than 1, which means that 12 pts was less legible than 8 pts. The effect of glare could have contributed to this small difference ratio in legibility, although this argument is not supported by the white-on-white for message II. However, these two cases further reinforces that a lack of contrast can affect legibility.

CHAPTER 6

SUMMARY AND RECOMMENDATIONS

One of the objectives of this research was to investigate whether the two instruments' readings gave the same results as the expert panel. The expert panel was used as the standard or reference control and therefore its reliability or internal consistency had to be proved before any correlation could be done. This evaluation showed a high experts' reliability (>.80) which meant that they could be used as a reference control against which the consumers' readings could be correlated. The results showed a very high correlation (>.90) between the experts' ratings and the readings obtained by the consumers using the two instruments. The correlations tended to be higher with the DMR visibility meter which had a positive linear relationship, than the polariscope which had a negative linear relationship. Both instruments gave results similar to the experts and therefore either one can be used to evaluate legibility of visual elements for packaging.

When compared against each other, the two instruments show a very high correlation (>.90) with negative slope. The negative relationship is explained by the reversed scales of the two instruments. Because it is hard to rate their performance without knowledge of their reliability, it is recommended that the two instruments be evaluated for reliability. With a known reliability it would then be possible to evaluate the two instruments' validity independently since the validity of a test is influenced by its reliability (27). There are some advantages offered by the polariscope which makes it superior to the DMR visibility meter.

The extended scale of the polariscope makes it more sensitive to changes which are not recorded in detail by the DMR visibility meter because of its compressed scale. One of the questions raised by this study is why the visual acuity of the observers did not have significant difference for the DMR visibility meter as it did for the polariscope. This can probably be attributed to the DMR visibility meter's lower sensitivity because of its compressed scale which is reflected by the closeness in the values obtained for legibility. This is also reinforced by the fact that visual acuity was almost significant as the results of the ANOVA show, which is also supported by the small ratio (1.1x) of the two acuities. On the other hand, an unpublished report on the use of polariscope for measuring legibility by Lockhart and Michel (8) showed clear differences between 20/20 and 20/30 visual acuities when testing for the legibility of visual elements on an ampoule and a vial. A general observation during this study was that subjects were more comfortable using the polariscope than the DMR visibility meter. This can be attributed to the polariscope's wide viewing screen, which was easier to manipulate. Subjects with bi-focals found it easier to use than the DMR visibility meter whose binocular viewing filters were small and restrictive.

With the results obtained with the polariscope, it is now easy to formulate a guideline that can be used by the manufacturers of CRP caps in testing legibility for the caps intended for consumers within the range of 20/20 to 20/30 vision.

Message, type size and color contrast affected the legibility of the CRP cap label copy. Results showed the two messages statistically significantly different for the experts and the consumers with both instruments. However, message effect, which is affected by message content and layout was not evaluated in this study although it was reflected in these results. A further study of message effect on legibility needs to be done.

Legibility was also affected by the type size. As the type got smaller, the message became hard to read. Legibility was severely diminished when the type was small (6 pts) and lacked contrast. The results showed a statistically significant difference between the black-on-white print and the white-on-white, ranging from one and a half to two times for the legibility ratio. A color contrast for these point sizes would greatly alleviate the current problem of legibility as this study has shown.

Alternatively the caps can be manufactured in any other color apart from the present white-on-white embossing, and then "reverse-printed" to offer a high contrast. It would be very hard to give an optimum type size for legibility purposes without an in depth research, but size 6 to 12pts were fairly legible, especially when colored black. Size of print will mainly depend on the size of the cap and the amount of information required.

Visual acuity of the observers affected the way they rated the legibility. The consumer group with 20/20 vision were able to read the two messages more easily than the 20/30 visual group. 30% of the two test groups had 20/30 vision the question that arises is whether tests for legibility should be conducted at 20/20. A study on legibility conducted by the School of Packaging for an age group between 56 and 76 years found a sizeable group with 20/30 and 20/40 visual acuity (See Appendix D, page 122). These two acuity levels should also be included in future legibility studies because they are so much a part of the age group reported to have difficulty with the readability of the CRP caps. This also raises the question of age effect on legibility which was not covered by this study. Note that in Tables 20 and 21 the regular decrease of legibility with increase in age in every visual acuity group suggests an age effect which, if tested statistically may prove to be significant.

Summary of Recommendations

1. Further studies of legibility on the elderly population because it was not included in this study. The message and age effect was not evaluated in this study, but its influence on legibility was observed. Further studies should be done to evaluate the effects.
2. During these studies, the order of presentation of the CRP caps to the test subjects should be fully randomized to avoid a learning effect which may affect the results.
3. Reverse-printing of the CRP caps by manufacturers for effective contrast to improve the legibility.
4. Further studies to evaluate the reliability and validity of the two instruments should be done in order to compare their performance.

LIST OF REFERENCES

REFERENCES

1. A.S.T.M., Specification No. D4267-83, "Standard Specification for Labels for Small-Volume (Less Than 100 Ml) Parenteral Drug Containers," (American Society for Testing and Materials), Philadelphia, Pa., 1983, pp. 1-2.

2. Bagaka, Joshua, G., Statistical Consultant, Research and Evaluation Services, Dept. of Urban Studies, interviewed by Mutune Wa Gitau, Michigan State University, E. Lansing, Michigan, 12.30 pm., June 13, 1988.

3. Barton, Leslie W., A Comparison of the Polariscope with the Tachistoscope for Evaluating Surface Designs of Consumer Packages, Thesis for the Degree of M.S., Michigan State University, 1968, pp. 4-32.

4. Bowdish, John, F., "Drug Labeling: A Challenge to ASTM.", ASTM Standardization News, August 1984, pp. 22-23.

5. General Radio Company, Polariscope, Cambridge, Mass., Nov. 1952, p. 2.

6. Glass, Gene, V., and Hopkins, Kenneth, D., "Split-Plot Design," Statistical Methods in Education and Psychology, Englewood Cliffs, N.J., Prentice-Hall, 1984, pp. 294-298.

7. Ibid., Table D.7.

8. Lockhart, Hugh, E., Dr., Professor, and Michel, Peg, Specialist, School of Packaging, Michigan State University, The Tachistoscope and Polariscope for Measuring Legibility, Report prepared for ASTM Committee D10, Denver, Colorado, October 12, 1983, p. 1.

9 Ibid., pp. 5.

10. Luckiesh, Matthew, "Meaning and Measurement of Visibility," Light, Vision and Seeing, D. Van Nostrand Co., New York, NY. 1944, pp. 177-193.

11. Ibid., pp. 139-157.

12. Luckiesh, M., and Moss, Frank, K., "Visibility of Objects," The Science of Seeing, D. Van Nostrand Co., New York, NY., 1937, pp. 168-180.

13. Morich, D., "Using the Tachistoscope, Semantic Differential and Preference Tests in Package Design Assessment," Handbook of Package Design Research, Wiley-Interscience, New York, NY., 1981, pp. 128-131.

14. Oborne, David, J., "Man-Man Communication: Words and Symbols," Ergonomics at Work, Wiley-Interscience, New York, NY., 1982, pp. 58.

15. Patterson, Donald, J., and Tinker, Miles, A., "Speed of Reading and Typography," How to Make Type Readable, Harper & Brothers Publishers, New York, 1940, pp. 1-12.

16. School of Packaging, Michigan State University, Innovative Child Resistant Packaging Systems, Final Report to Consumer Product Safety Commission (CPSC), April 28, 1988, pp. 38-39.

17. Swope, Lee, "Visionmetric Testing: How a Package Communicates," Handbook of Package Design Research, Wiley-Interscience, New York, NY., 1981, pp. 142-143.

18. "The Eyes Have It," Modern Packaging, Vol. 34, No.2, October, 1960, pp. 124-125.

19. Theodore, Horvath, "Product-Moment Correlation," Basic Statistics for Behavioral Sciences, Little, Brown and Company, Boston, Toronto, 1985, pp. 262-264.

20. Ibid., pp. 360.

21. Ibid., pp. 223-225.

22. Tinker, Miles. A., Legibility of Print, The Iowa State University Press, Ames, Iowa, 1963, pp. 7-8.

23. Ibid., pp. 9-31.

24. Tobolski, Francis, P., Carol Stream, Illinois, DMR Visibility Meter, Letter sent to Lockhart, Hugh, E., Dr., Professor, School of Packaging, Michigan State University, August 25, 1987.

25. United States Pharmacopeia (USP), Twenty-First Revision, Washington DC: The United States Pharmacopeial Convention, 1985, p. 24.

26. Webster's Third New International Dictionary of the English Language Unabridged, Springfield, Mass., G & C. Merrian Co., 1976.

27. William, Mehrens, A., and Lehmann, Irvin, J., "Reliability and Validity," Measurement and Evaluation in Education and Psychology, Holt, Rinehart and Winston, New York, Chicago, 1975, pp. 87-109.

APPENDIX A:

EXPERT PANEL DATA

Raw Data: Legibility Rating by the Expert Panel

				Cap No. and Rating on a Scale of 1-10															
No.	Sex	Age	Visual Acuity	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	M	40	20/20	8	1	7	4	5	3	7	3	8	2	7	4	6	3	7	4
2	M	40	20/20	9	2	5	8	4	6	7	5	9	2	7	8	4	6	7	6
3	M	60	20/30	8	1	4	7	2	3	5	3	8	1	5	6	2	4	5	3
4	M	32	20/20	10	5	10	10	4	9	10	5	10	2	9	10	5	9	10	9
5	M	42	20/20	10	3	7	9	3	5	6	5	10	3	6	7	3	5	8	5
6	F	55	20/20	10	1	4	9	2	7	9	1	10	1	4	9	2	8	4	1
7	F	45	20/20	9	3	5	7	3	7	7	4	9	3	5	7	4	6	7	4
8	M	41	20/20	9	3	5	8	3	5	7	4	9	4	5	7	3	5	5	3
9	M	44	20/20	10	1	6	7	3	5	7	4	10	1	6	6	4	5	6	4
10	M	37	20/20	8	2	5	9	2	7	8	4	10	1	4	8	3	6	8	3

Mean Rating of Legibility of 8 Caps by the Expert Panel

Experts	Cap No. and Rating*							
No.	1	2	3	4	5	6	7	8
1	8.0	1.5	7.0	4.0	5.5	3.0	7.0	3.5
2	9.0	2.0	6.0	8.0	4.0	6.0	7.0	5.5
3	8.0	1.0	4.5	6.5	2.0	3.5	5.0	3.0
4	10.0	3.5	9.5	10.0	4.5	9.0	10.0	7.0
5	10.0	3.0	6.5	8.0	3.0	5.0	7.0	5.0
6	10.0	1.0	4.0	9.0	2.0	7.5	6.5	1.0
7	9.0	3.5	5.0	7.0	3.5	6.5	7.0	4.0
8	9.0	3.5	5.0	7.5	3.0	5.0	6.0	3.5
9	10.0	1.0	6.0	6.5	3.5	5.0	6.5	4.0
10	9.0	1.5	4.5	8.5	2.5	6.5	8.0	3.5
Mean	9.0	2.0	6.0	8.0	4.0	6.0	7.0	4.0
Std. Dev.	0.9	1.1	1.6	1.7	1.3	1.8	1.5	1.6

* Caps 1, 3, 5 and 7 = Message I

Caps 2, 4, 6 and 8 = Message II

Experts Reliability Analysis - Scale (Rating)
(a) Reliability Coefficient Matrix
(Reported in Standardized Item Alpha)

	Expert									
Expert	1	2	3	4	5	6	7	8	9	10
1	-									
2	.5362	-								
3	.5336	.8025	-							
4	.7028	.7265	.2864	-						
5	-.0259	.3840	.7718	-1.6397	-					
6	.6582	.7897	.7718	.8549	-.2236	-				
7	.5921	.6115	.6282	.6599	.2362	.9045	-			
8	.2204	.7845	.6914	.3955	.5526	.3479	-.1959	-		
9	.8402	.5169	.4226	.8689	-.9076	.8287	.7346	.2988	-	
10	.2149	.7183	.2970	.0000	.4825	.0000	-.3492	.6002	-.2937	-

Overall Reliability Coefficient - 10 items : Alpha = .8295 Standardized Item Alpha = .8278

Experts Reliability Analysis - Scale (Rating)
(b) Correlation Matrix

Expert	1	2	3	4	5	6	7	8	9	10
1	1.0000									
2	.3663	1.0000								
3	.3639	.6701	1.0000							
4	.5417	.5705	.1671	1.0000						
5	-.0128	.2376	.6284	-.4505	1.0000					
6	.4905	.6524	.3374	.7465	-.1006	1.0000				
7	.4206	.4404	.4580	.4924	.1339	.8256	1.0000			
8	.1239	.6454	.5283	.2465	.3818	.2106	-.0892	1.0000		
9	.7244	.3485	.2679	.7682	-.3122	.7075	.5805	.1757	1.0000	
10	.1204	.5604	.1744	.0000	.3180	.0000	-.1487	.4287	-.1280	1.0000

SAS
ANALYSIS OF VARIANCE PROCEDURE FOR THE EXPERTS

DEPENDENT VARIABLE: SCALE

Source:	DF	SUM OF SQUARES	MEAN SQUARE
Model:	7	380.98125000	54.42589286
Error:	72	143.40625000	1.99175347
Corrected Total:	79	524.38750000	

F VALUE	PR>F	R-SQUARE	C.V.
27.33	0.0001	0.726526	25.2581

ROOT MSE	SCALE MEAN
1.41129496	5.58750000

SOURCE:	DF	ANOVA SS	F VALUE	PR>F
MESSAGE	1	45.00000000	29.59	0.0001
COLOR	1	248.51250000	124.77	0.0001
SIZE	2	129.05625000	23.40	0.0001
MESSAGE*COLOR	1	0.00000000	0.00	1.0000
MESSAGE*SIZE	2	0.00000000	.	.
COLOR*SIZE	2	0.10625000	0.03	0.9737

SAS
ANALYSIS OF VARIANCE PROCEDURE FOR THE EXPERTS
MAIN EFFECTS TEST

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
MESSAGE	2	1 2
COLOR	2	1 2
SIZE	3	1 2 3

NUMBER OF OBSERVATIONS IN DATA SET = 80

SAS

ANALYSIS OF VARIANCE PROCEDURE FOR THE EXPERTS

MAIN EFFECTS TEST

TUKEY'S STUDENTIZED RANGE (HSD) TEST FOR VARIABLE: SCALE
NOTE: THIS TEST CONTROLS THE TYPE I EXPERIMENTWISE ERROR RATE, BUT GENERALLY HAS A HIGHER TYPE II ERROR RATE THAN REGWQ

ALPHA=0.05 DF=72 MSE=1.99175
CRITICAL VALUE OF STUDENTIZED RANGE=2.819
MINIMUM SIGNIFICANT DIFFERENCE=.62909

MEANS WITH THE SAME LETTER ARE NOT SIGNIFICANTLY DIFFERENT.

TUKEY GROUPING	MEAN	N	MESSAGE
A	6.3375	40	1
B	4.8375	40	2

SAS

ANALYSIS OF VARIANCE PROCEDURE FOR THE EXPERTS

MAIN EFFECTS TEST

TUKEY'S STUDENTIZED RANGE (HSD) TEST FOR VARIABLE: SCALE
NOTE: THIS TEST CONTROLS THE TYPE I EXPERIMENTWISE ERROR RATE, BUT GENERALLY HAS A HIGHER TYPE II ERROR RATE THAN REGWQ

ALPHA=0.05 DF=72 MSE=1.99175
CRITICAL VALUE OF STUDENTIZED RANGE=2.819
MINIMUM SIGNIFICANT DIFFERENCE=.62909

MEANS WITH THE SAME LETTER ARE NOT SIGNIFICANTLY DIFFERENT.

TUKEY GROUPING	MEAN	N	COLOR
A	7.3500	40	1
B	3.8250	40	2

SAS

ANALYSIS OF VARIANCE PROCEDURE FOR THE EXPERTS

MAIN EFFECTS TEST

TUKEY'S STUDENTIZED RANGE (HSD) TEST FOR VARIABLE: SCALE
NOTE: THIS TEST CONTROLS THE TYPE I EXPERIMENTWISE ERROR RATE

ALPHA=0.05 CONFIDENCE=0.95 DF=72 MSE=1.99175
CRITICAL VALUE OF STUDENTIZED RANGE=3.384

COMPARISONS SIGNIFICANT AT THE 0.05 LEVEL ARE INDICATED BY '***'

SIZE COMPARISON		SIMULTANEOUS LOWER CONFIDENCE LIMIT	DIFFERENCE BETWEEN MEANS	SIMULTANEOUS UPPER CONFIDENCE LIMIT	
3	-2	1.1126	2.0375	2.9624	***
3	-1	2.5070	3.5750	4.6430	***
2	-3	-2.9624	-2.0375	-1.1126	***
2	-1	0.6126	1.5375	2.4624	***
1	-3	-4.6430	-3.5750	-2.5070	***
1	-2	-2.4624	-1.5375	-0.6126	***

An Example of a Mathematical Computation of Tukey's HSD for Color Contrast for the Expert Panel.

$$HSD = q_{\alpha,v} \sqrt{\frac{MSE}{n}}$$

α = 0.05

v = 72

$q_{\alpha,v}$ = 2.819 > given in Appendix A, page 105

MSE = 1.99175

n = 40

$$HSD = 2.819 \sqrt{\frac{1.99175}{40}}$$

$$= .6290$$

APPENDIX B:

CONSUMER DATA

CONSUMER DATA

RAW DATA: LEGIBILITY READINGS BY THE CONSUMERS USING POLARISCOPE AND DMR VISIBILITY METER

				POLARISCOPE								DMR VISIBILITY METER							
				Cap No.* & Degrees of Rotation								Cap No. & Visual Angle in Minutes							
No.	Age Group	Sex	Visual Acuity	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
1	18-25 yrs	F	20/30	6	19	13	8	8	11	9	10	20	11	14	20	13	16	20	15
2	18-25 yrs	F	20/20	3	12	17	7	5	10	7	11	14	8	11	15	12	11	16	12
3	18-25 yrs	M	20/20	4	11	5	5	7	6	5	6	20	15	18	20	15	19	20	16
4	18-25 yrs	F	20/30	3	9	5	5	14	7	5	10	20	12	18	20	13	17	18	14
5	18-25 yrs	F	20/20	3	9	4	4	5	5	4	5	20	14	18	20	15	19	20	18
6	18-25 yrs	M	20/20	5	12	10	6	7	7	4	8	20	11	17	17	11	13	17	15
7	18-25 yrs	F	20/20	4	11	7	6	8	8	7	7	17	11	15	17	12	12	13	12
8	18-25 yrs	F	20/20	5	12	9	6	8	7	6	10	20	9	11	15	10	12	13	11
9	18-25 yrs	F	20/20	5	11	5	4	9	8	5	7	15	8	13	13	11	11	14	11
10	18-25 yrs	M	20/30	4	17	8	7	13	9	6	5	20	8	12	12	10	10	12	10
11	30-54 yrs	F	20/20	4	18	9	7	11	10	9	12	16	6	11	11	8	10	11	9
12	30-54 yrs	F	20/20	5	18	11	10	13	11	7	14	20	2	6	11	4	11	11	8
13	30-54 yrs	M	20/30	17	42	26	22	35	28	22	32	13	4	5	11	3	7	8	4
14	30-54 yrs	F	20/20	5	16	9	8	10	9	8	10	20	9	13	20	11	15	20	14
15	30-54 yrs	M	20/20	7	21	12	9	17	11	8	17	20	5	11	16	7	11	14	9
16	30-54 yrs	F	20/30	5	13	7	5	10	8	4	10	20	9	15	15	6	13	15	12
17	30-54 yrs	M	20/20	4	17	6	3	9	4	5	7	20	14	14	20	13	16	20	14
18	30-54 yrs	M	20/20	3	13	7	6	10	8	7	10	20	11	17	20	14	20	20	17
19	30-54 yrs	F	20/30	10	40	15	16	20	23	15	19	20	8	10	15	11	12	17	10
20	30-54 yrs	M	20/20	5	25	7	5	8	7	6	9	20	9	14	20	9	13	20	12

* Caps No. 1, 3, 5 and 7 = Message I and Caps No. 2, 4, 6 and 8 = Message II

LEGIBILITY MEAN READINGS * FOR 20/20'S AND 20/30'S CONSUMERS USING POLARISCOPE

	20/20		20/30	
Cap No.	Mean	STD. DEV.	Mean	STD. DEV.
1	05	1.1	08	3.8
2	14	1.5	24	10.8
3	09	3.4	13	6.8
4	06	1.8	11	5.1
5	09	1.9	17	7.9
6	08	2.1	15	6.2
7	06	1.4	11	5.6
8	10	2.8	14	7.0

* The mean readings have been pooled for the two Age groups.

LEGIBILITY MEAN READINGS* FOR 20/20'S AND 20/30'S CONSUMERS USING DMR VISIBILITY METER

	20/20		20/30	
Cap No.	Mean	STD. DEV.	Mean	STD. DEV.
1	19	2.1	19	2.1
2	10	3.4	09	2.4
3	14	3.3	13	4.1
4	17	3.1	16	3.5
5	11	2.8	10	2.9
6	14	3.6	13	3.5
7	17	3.7	15	4.5
8	13	3.1	11	3.4

* The mean readings have been pooled for the two Age groups.

LAYOUT OR BLOCK DIAGRAM FOR SPF-pru.q (SPF-222.4) DESIGN USED FOR ANOVA IN THIS RESEARCH

A	MESSAGE I				MESSAGE II			
C	B/W		W/W		B/W		W/W	
D	8pts	12pts	8pts	12pts	6pts	8pts	6pts	8pts
B	S1	S1	S1	S1	S1	S1	S1	S1
	S2	S2	S2	S2	S2	S2	S2	S2
	S3	S3	S3	S3	S3	S3	S3	S3
	S4	S4	S4	S4	S4	S4	S4	S4
	S5	S5	S5	S5	S5	S5	S5	S5
	S6	S6	S6	S6	S6	S6	S6	S6
20/20	S7	S7	S7	S7	S7	S7	S7	S7
	S8	S8	S8	S8	S8	S8	S8	S8
	S9	S9	S9	S9	S9	S9	S9	S9
	S10	S10	S10	S10	S10	S10	S10	S10
	S11	S11	S11	S11	S11	S11	S11	S11
	S12	S12	S12	S12	S12	S12	S12	S12
	S13	S13	S13	S13	S13	S13	S13	S13
	S14	S14	S14	S14	S14	S14	S14	S14
	S1	S1	S1	S1	S1	S1	S1	S1
	S2	S2	S2	S2	S2	S2	S2	S2
20/30	S3	S3	S3	S3	S3	S3	S3	S3
	S4	S4	S4	S4	S4	S4	S4	S4
	S5	S5	S5	S5	S5	S5	S5	S5
	S6	S6	S6	S6	S6	S6	S6	S6

Treatment Levels: A - Message
C - Color
D - Size
B - Visual Acuity

SPF - ACD.B

Structural Model:

$$X_{ijkl} = \mu+\alpha_i+y_k+\delta_l+\beta_j+(\alpha\delta)_{ik}+(\alpha\delta)_{il}+(\alpha\beta)_{ij}+(\alpha y\delta)_{ikl}+(\alpha y\beta)_{ikj}+(\alpha\delta\beta)_{ilj}+(\alpha y\delta\beta)_{iklj}+(y\delta)_{kl}+(y\beta)_{kj}+(y\delta\beta)_{klj}+(\delta\beta)_{lj}+\epsilon_o(ijkl)$$

where μ = Grand mean of treatment populations.

α_i = Effect of treatment i of message, which is constant for all subjects within treatment population i.

y_k = Effect of treatment k of visual acuity.

δ_l = Effect of treatment l of color.

β_j = Effect of treatment j of size.

$\epsilon_o(ijkl)$ = Experimental error, which is independent of other E's and is normally distributed with mean = 0 and variance = σ_ϵ^2.

An Example of Mathematical Computation of Correlation (r) for Experts Rating of Legibility vs. Consumers 20/20 Readings Using the Polariscope

Variable X: Experts' rating of Cap legibility in mean order ranking.

Variable Y: Consumers 20/20 readings of legibility by polariscope in mean order ranking.

Cap No.	X	Y	X^2	Y^2	XY
5	4	9	16	81	36
3	6	9	36	81	54
7	7	6	49	36	42
1	9	5	81	25	45
2	2	14	4	196	28
8	4	10	16	100	40
6	6	8	36	64	48
4	8	6	64	36	48
	ΣX46	ΣY67	ΣX^2 302	ΣY^2 619	ΣXY341

N = 8

$$r = \frac{N\Sigma XY - \Sigma X \Sigma Y}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2]\,[N\Sigma Y^2 - (\Sigma Y)^2]}}$$

$$= \frac{(8)(341) - (46)(67)}{\sqrt{[(8)(302) - (46)^2]\,[(8)(618) - (67)^2]}}$$

$$= \frac{-\ 354}{372.7}$$

$$= -\ 0.9496$$

$$= -0.950$$

SAS

ANALYSIS OF VARIANCE PROCEDURE FOR CONSUMERS USING POLARISCOPE

DEPENDENT VARIABLE: SCALE

Source:	DF	SUM OF SQUARES	MEAN SQUARE
Model:	15	2892.38660714	192.82577381
Error:	144	4134.35714286	28.71081349
Corrected Total:	159	7026.74375000	

F VALUE	PR>F	R-SQUARE	C.V.
6.72	0.0001	0.411625	54.2264

ROOT MSE	SCALE MEAN
5.35824724	9.88125000

SOURCE:	DF	ANOVA SS	F VALUE	PR>F
MESSAGE	1	333.50625000	11.62	0.0008
VISUAL	1	971.80029762	33.85	0.0001
COLOR	1	888.30625000	30.94	0.0001
SIZE	2	810.50625000	14.11	0.0001
MESSAGE*VISUAL	1	19.65744048	0.68	0.4094
MESSAGE*COLOR	1	16.25625000	0.57	0.4530
MESSAGE*SIZE	2	0.00000000	.	.
MESSAGE*VISUAL*COLOR	1	1.18125000	0.04	0.8395
MESSAGE*VISUAL*SIZE	1	0.00000000	.	.
MESSAGE*COLOR*SIZE	0	0.00000000	.	.
MESSAGE*VISU*COLO*SIZE	1	7.53363095	.	.
VISUAL*COLOR	1	30.28601190	1.05	0.3061
VISUAL*SIZE	2	68.50267857	1.19	0.3063
VISUAL*COLOR*SIZE	2	3.18363095	0.06	0.9461
COLOR*SIZE	2	102.95625000	1.79	0.1702

SAS

ANALYSIS OF VARIANCE PROCEDURE FOR CONSUMERS USING POLARISCOPE-MAIN EFFECTS TEST

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
MESSAGE	2	1 2
VISUAL ACUITY	2	1 2
COLOR	2	1 2
SIZE	3	1 2 3

NUMBER OF OBSERVATIONS IN DATA SET = 160

SAS

ANALYSIS OF VARIANCE PROCEDURE FOR CONSUMERS USING POLARISCOPE - MAIN EFFECTS TEST

TUKEY'S STUDENTIZED RANGE (HSD) TEST FOR VARIABLE: SCALE

NOTE: THIS TEST CONTROLS THE TYPE I EXPERIMENTWISE ERROR RATE, BUT GENERALLY HAS A HIGHER TYPE II ERROR RATE THAN REGWQ

ALPHA=0.05 DF=144 MSE=28.7108
CRITICAL VALUE OF STUDENTIZED RANGE=2.795
MINIMUM SIGNIFICANT DIFFERENCE=1.6746

MEANS WITH THE SAME LETTER ARE NOT SIGNIFICANTLY DIFFERENT.

TUKEY GROUPING	MEAN	N	MESSAGE
A	11.3250	80	2
B	8.4375	80	1

SAS

ANALYSIS OF VARIANCE PROCEDURE FOR CONSUMERS USING POLARISCOPE - MAIN EFFECTS TEST

TUKEY'S STUDENTIZED RANGE (HSD) TEST FOR VARIABLE: SCALE
NOTE: THIS TEST CONTROLS THE TYPE I EXPERIMENTWISE ERROR RATE, BUT GENERALLY HAS A HIGHER TYPE II ERROR RATE THAN REGWQ

ALPHA=0.05 DF=144 MSE=28.7108
CRITICAL VALUE OF STUDENTIZED RANGE=2.795
MINIMUM SIGNIFICANT DIFFERENCE=1.8271

WARNING: CELL SIZES ARE NOT EQUAL.
HARMONIC MEAN OF CELL SIZES=67.2

MEANS WITH THE SAME LETTER ARE NOT SIGNIFICANTLY DIFFERENT.

TUKEY GROUPING	MEAN	N	VISUAL ACUITY
A	13.6458	48	2
B	8.2679	112	1

SAS

ANALYSIS OF VARIANCE PROCEDURE FOR CONSUMERS USING POLARISCOPE - MAIN EFFECTS TEST

TUKEY'S STUDENTIZED RANGE (HSD) TEST FOR VARIABLE: SCALE
NOTE: THIS TEST CONTROLS THE TYPE I EXPERIMENTWISE ERROR RATE, BUT GENERALLY HAS A HIGHER TYPE II ERROR RATE THAN REGWQ

ALPHA=0.05 DF=144 MSE=28.7108
CRITICAL VALUE OF STUDENTIZED RANGE=2.795
MINIMUM SIGNIFICANT DIFFERENCE=1.6746

MEANS WITH THE SAME LETTER ARE NOT SIGNIFICANTLY DIFFERENT.

TUKEY GROUPING	MEAN	N	COLOR
A	12.2375	80	2
B	7.5250	80	1

SAS

ANALYSIS OF VARIANCE PROCEDURE FOR CONSUMERS USING POLARISCOPE - MAIN EFFECTS TEST

TUKEY'S STUDENTIZED RANGE (HSD) TEST FOR VARIABLE: SCALE

NOTE: THIS TEST CONTROLS THE TYPE I EXPERIMENTWISE ERROR RATE

ALPHA=0.05 CONFIDENCE=0.95 DF=144 MSE=28.7108

CRITICAL VALUE OF STUDENTIZED RANGE=3.349

COMPARISONS SIGNIFICANT AT THE 0.05 LEVEL ARE INDICATED BY '***'

SIZE COMPARISON		SIMULTANEOUS LOWER CONFIDENCE LIMIT	DIFFERENCE BETWEEN MEANS	SIMULTANEOUS UPPER CONFIDENCE LIMIT	
1	-2	1.8802	4.3375	6.7948	***
1	-3	3.2625	6.1000	8.9375	***
2	-1	-6.7948	-4.3375	-1.8802	***
2	-3	-0.6948	1.7625	4.2198	
3	-1	-8.9375	-6.1000	-3.2625	***
3	-2	-4.2198	-1.7625	0.6948	

SAS

ANALYSIS OF VARIANCE PROCEDURE FOR CONSUMERS USING DMR VISILITY METER

DEPENDENT VARIABLE: SCALE

Source:	DF	SUM OF SQUARES	MEAN SQUARE
Model:	15	1490.14761905	99.34317460
Error:	144	1684.95238095	11.70105820
Corrected Total:	159	3175.10000000	

F VALUE	PR>F	R-SQUARE	C.V.
8.49	0.0001	0.469323	25.0141

ROOT MSE	SCALE MEAN
3.42068096	13.67500000

SOURCE:	DF	ANOVA SS	F VALUE	PR>F
MESSAGE	1	126.02500000	10.77	0.0013
VISUAL	1	43.88571429	3.75	0.0547
COLOR	2	960.40000000	82.08	0.0001
SIZE	1	432.85000000	18.50	0.0001
MESSAGE*VISUAL	1	0.86785714	0.07	0.7858
MESSAGE*COLOR	2	18.22500000	1.56	0.2140
MESSAGE*SIZE	2	0.00000000	.	.
MESSAGE*VISUAL*COLOR	0	1.00119048	0.09	0.7703
MESSAGE*VISUAL*SIZE	0	0.00000000	.	.
MESSAGE*COLOR*SIZE	0	0.00000000	.	.
MESSAGE*VISU*COLO*SIZE	0	0.00000000	.	.
VISUAL*COLOR	1	1.21904762	0.10	0.7473
VISUAL*SIZE	2	5.62619048	0.24	0.7866
VISUAL*COLOR*SIZE	2	3.43809524	0.15	0.8635
COLOR*SIZE	2	9.80000000	0.42	0.6587

SAS

ANALYSIS OF VARIANCE PROCEDURE FOR CONSUMERS USING
DMR VISIBILITY METER - MAIN EFFECTS TEST

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
MESSAGE	2	1 2
VISUAL ACUITY	2	1 2
COLOR	2	1 2
SIZE	3	1 2 3

NUMBER OF OBSERVATIONS IN DATA SET = 160

SAS

ANALYSIS OF VARIANCE PROCEDURE FOR CONSUMERS USING
DMR VISIBILITY METER - MAIN EFFECTS TEST

TUKEY'S STUDENTIZED RANGE (HSD) TEST FOR VARIABLE: SCALE
NOTE: THIS TEST CONTROLS THE TYPE I EXPERIMENTWISE ERROR RATE, BUT GENERALLY HAS A HIGHER TYPE II ERROR RATE THAN REGWQ

ALPHA=0.05 DF=144 MSE=11.7095
CRITICAL VALUE OF STUDENTIZED RANGE=2.795
MINIMUM SIGNIFICANT DIFFERENCE=1.069

MEANS WITH THE SAME LETTER ARE NOT SIGNIFICANTLY DIFFERENT.

TUKEY GROUPING	MEAN	N	MESSAGE
A	14.5625	80	1
B	12.7875	80	2

SAS

ANALYSIS OF VARIANCE PROCEDURE FOR CONSUMERS USING DMR VISIBILITY METER - MAIN EFFECTS TEST

TUKEY'S STUDENTIZED RANGE (HSD) TEST FOR VARIABLE: SCALE
NOTE: THIS TEST CONTROLS THE TYPE I EXPERIMENTWISE ERROR RATE, BUT GENERALLY HAS A HIGHER TYPE II ERROR RATE THAN REGWQ

ALPHA=0.05 DF=144 MSE=11.7011
CRITICAL VALUE OF STUDENTIZED RANGE=2.795
MINIMUM SIGNIFICANT DIFFERENCE=1.1664

WARNING: CELL SIZES ARE NOT EQUAL.
HARMONIC MEAN OF CELL SIZES=67.2

MEANS WITH THE SAME LETTER ARE NOT SIGNIFICANTLY DIFFERENT.

TUKEY GROUPING	MEAN	N	VISUAL ACUITY
A	14.0179	112	1
A	12.8750	48	2

SAS

ANALYSIS OF VARIANCE PROCEDURE FOR CONSUMERS USING DMR VISIBILITY METER - MAIN EFFECTS TEST

TUKEY'S STUDENTIZED RANGE (HSD) TEST FOR VARIABLE: SCALE
NOTE: THIS TEST CONTROLS THE TYPE I EXPERIMENTWISE ERROR RATE, BUT GENERALLY HAS A HIGHER TYPE II ERROR RATE THAN REGWQ

ALPHA=0.05 DF=144 MSE=11.7011
CRITICAL VALUE OF STUDENTIZED RANGE=2.795
MINIMUM SIGNIFICANT DIFFERENCE=1.069

MEANS WITH THE SAME LETTER ARE NOT SIGNIFICANTLY DIFFERENT.

TUKEY GROUPING	MEAN	N	COLOR
A	16.1250	80	1
B	11.2250	80	2

SAS

ANALYSIS OF VARIANCE PROCEDURE FOR CONSUMERS USING
DMR VISIBILITY METER - MAIN EFFECTS TEST

TUKEY'S STUDENTIZED RANGE (HSD) TEST FOR VARIABLE: SCALE
NOTE: THIS TEST CONTROLS THE TYPE I EXPERIMENTWISE ERROR RATE

ALPHA=0.05 CONFIDENCE=0.95 DF=144 MSE=11.7011
CRITICAL VALUE OF STUDENTIZED RANGE=3.349

COMPARISONS SIGNIFICANT AT THE 0.05 LEVEL ARE INDICATED BY '***'

SIZE COMPARISON		SIMULTANEOUS LOWER CONFIDENCE LIMIT	DIFFERENCE BETWEEN MEANS	SIMULTANEOUS UPPER CONFIDENCE LIMIT	
3	-2	0.6563	2.2250	3.7937	***
3	-1	2.8386	4.6500	6.4614	***
2	-3	-3.7937	-2.2250	-0.6563	***
2	-1	0.8563	2.4250	3.9937	***
1	-3	-6.4614	-4.6500	-2.8386	***
1	-2	-3.9937	-2.4250	-0.8563	***

APPENDIX C:

PROGRAMS FOR COMPUTING RELIABILITY AND ANOVA

SPSS PROGRAM FOR COMPUTING THE EXPERT PANEL'S RELIABILITY.

```
0  FILE HANDLE MWG2/NAME='RELIAB DAT A'
0  DATA LIST FILE=MWG2
0  /CAPID 1-2 EXPERT1 TO EXPERT10 3-22
```

THE ABOVE DATA LIST STATEMENT WILL READ 1 RECORDS FROM FILE MWG2

VARIABLE	REC	START	END	FORMAT	WIDTH	DEC
CAPID	1	1	2	F	2	0
EXPERT1	1	3	4	F	2	0
EXPERT2	1	5	6	F	2	0
EXPERT3	1	7	8	F	2	0
EXPERT4	1	9	10	F	2	0
EXPERT5	1	11	12	F	2	0
EXPERT6	1	13	14	F	2	0
EXPERT7	1	15	16	F	2	0
EXPERT8	1	17	18	F	2	0
EXPERT9	1	19	20	F	2	0
EXPERT10	1	21	22	F	2	0

END OF DATALIST TABLE.

```
0   RELIABILITY VARIABLES=EXPERT1 TO EXPERT10/
0      SCALE(RATING1)=EXPERT1 EXPERT2/
0      SCALE(RATING2)=EXPERT1 EXPERT3/
0      SCALE(RATING3)=EXPERT1 EXPERT4/
0      SCALE(RATING4)=EXPERT1 EXPERT5/
0      SCALE(RATING5)=EXPERT1 EXPERT6/
0      SCALE(RATING6)=EXPERT1 EXPERT7/
0      SCALE(RATING7)=EXPERT1 EXPERT8/
0      SCALE(RATING8)=EXPERT1 EXPERT9/
0      SCALE(RATING9)=EXPERT1 EXPERT10/
0      SCALE(RATING10)=EXPERT2 EXPERT3/
0      SCALE(RATING11)=EXPERT2 EXPERT4/
0      SCALE(RATING12)=EXPERT2 EXPERT5/
0      SCALE(RATING13)=EXPERT2 EXPERT6/
0      SCALE(RATING14)=EXPERT2 EXPERT7/
0      SCALE(RATING15)=EXPERT2 EXPERT8/
0      SCALE(RATING16)=EXPERT2 EXPERT9/
0      SCALE(RATING17)=EXPERT2 EXPERT10/
0      SCALE(RATING18)=EXPERT3 EXPERT4/
0      SCALE(RATING19)=EXPERT3 EXPERT5/
0      SCALE(RATING20)=EXPERT3 EXPERT5/
0      SCALE(RATING21)=EXPERT3 EXPERT7/
0      SCALE(RATING22)=EXPERT3 EXPERT8/
0      SCALE(RATING23)=EXPERT3 EXPERT9/
0      SCALE(RATING24)=EXPERT3 EXPERT10/
0      SCALE(RATING25)=EXPERT4 EXPERT5/
0      SCALE(RATING26)=EXPERT4 EXPERT6/
0      SCALE(RATING27)=EXPERT4 EXPERT7/
0      SCALE(RATING28)=EXPERT4 EXPERT8/
0      SCALE(RATING29)=EXPERT4 EXPERT9/
```

```
0      SCALE(RATING30)=EXPERT4 EXPERT10/
0      SCALE(RATING31)=EXPERT5 EXPERT6/
0      SCALE(RATING32)=EXPERT5 EXPERT7/
0      SCALE(RATING33)=EXPERT5 EXPERT8/
0      SCALE(RATING34)=EXPERT5 EXPERT9/
0      SCALE(RATING35)=EXPERT5 EXPERT10/
0      SCALE(RATING36)=EXPERT6 EXPERT7/
0      SCALE(RATING37)=EXPERT6 EXPERT8/
0      SCALE(RATING38)=EXPERT6 EXPERT9/
0      SCALE(RATING39)=EXPERT6 EXPERT10/
0      SCALE(RATING40)=EXPERT7 EXPERT8/
0      SCALE(RATING41)=EXPERT7 EXPERT9/
0      SCALE(RATING42)=EXPERT7 EXPERT10/
0      SCALE(RATING43)=EXPERT8 EXPERT9/
0      SCALE(RATING44)=EXPERT8 EXPERT10/
0      SCALE(RATING45)=EXPERT9 EXPERT10/
0      SCALE(RATING)=EXPERT1 TO EXPERT10/
0      MODEL=ALPHA
0  STATISTICS 3

****** METHOD 2 (COVARIANCE MATRIX) WILL BE USED FOR THIS ANALYSIS ******
```

SAS ANOVA PROGRAM FOR MACHINE I (P-SCOPE), MACHINE 2 (DMR METER) AND THE EXPERT PANEL

```
CMS FILEDEP GITAU DISK CONSUMER DATA A;
DATA ALL;
   INFILE GITAU;
   INPUT ID 1-2 CAPID 3 MACHINE 4 SCALE 6-7 AGE 9 VISUAL 11;
COLOR = 1;
 IF CAPID EQ 2 OR CAPID EQ 3 OR CAPID EQ 5 OR CAPID EQ 8 THEN COLOR = 2;
MESSAGE = 1;
 IF CAPID EQ 2 OR CAPID EQ 4 OR CAPID EQ 6 OR CAPID EQ 8 THEN MESSAGE=2;
SIZE = 1;
 IF CAPID EQ 4 OR CAPID EQ 5 OR CAPID EQ 7 OR CAPID EQ 8 THEN SIZE =2;
 IF CAPID EQ 1 OR CAPID EQ 3 THEN SIZE = 3;
DATA MACHINE1; SET ALL;
  IF MACHINE = 1;
PROC ANOVA;
   CLASS MESSAGE VISUAL COLOR SIZE;
   MODEL SCALE = MESSAGE VISUAL COLOR SIZE
         MESSAGE*VISUAL MESSAGE*COLOR MESSAGE*SIZE
         MESSAGE*VISUAL*COLOR MESSAGE*VISUAL*SIZE
         MESSAGE*COLOR*SIZE MESSAGE*VISUAL*COLOR*SIZE
         VISUAL*COLOR VISUAL*SIZE VISUAL*COLOR*SIZE
         COLOR*SIZE;
   MEANS MESSAGE VISUAL COLOR SIZE
         MESSAGE*VISUAL MESSAGE*COLOR* MESSAGE*SIZE
         MESSAGE*VISUAL*COLOR MESSAGE*VISUAL*SIZE
         MESSAGE*COLOR*SIZE MESSAGE*VISUAL*COLOR*SIZE
         VISUAL*COLOR VISUAL*SIZE VISUAL*COLOR*SIZE
         COLOR*SIZE / TUKEY;
DATA MACHINE2; SET ALL;
  IF MACHINE = 2;
PROC ANOVA;
   CLASS MESSAGE VISUAL COLOR SIZE;
   MODEL SCALE = MESSAGE VISUAL COLOR SIZE
         MESSAGE*VISUAL MESSAGE*COLOR MESSAGE*SIZE
         MESSAGE*VISUAL*COLOR MESSAGE*VISUAL*SIZE
         MESSAGE*COLOR*SIZE MESSAGE*VISUAL*COLOR*SIZE
         VISUAL*COLOR VISUAL*SIZE VISUAL*COLOR*SIZE
         COLOR*SIZE;
   MEANS MESSAGE VISUAL COLOR SIZE
         MESSAGE*VISUAL MESSAGE*COLOR MESSAGE*SIZE
         MESSAGE*VISUAL*COLOR MESSAGE*VISUAL*SIZE
         MESSAGE*COLOR*SIZE MESSAGE*VISUAL*COLOR*SIZE
         VISUAL*COLOR VISUAL*SIZE VISUAL*COLOR*SIZE
         COLOR*SIZE / TUKEY;
NOTE: INFILE GITAU IS FILE EXPERT DATA A1
    PROC ANOVA:
       CLASS MESSAGE COLOR SIZE
       MODEL SCALE = MESSAGE COLOR SIZE
             MESSAGE*COLOR MESSAGE*SIZE COLOR*SIZE
             MESSAGE*COLOR SIZE;
       MEANS MESSAGE COLOR SIZE
             MESSAGE*COLOR MESSAGE*SIZE COLOR*SIZE
             MESSAGE*COLOR SIZE /TUKEY;
```

APPENDIX D:

LEGIBILITY TEST DATA FOR 56-76+ CONSUMER AGE GROUP USING THE POLARISCOPE

LEGIBILITY TEST FOR 56-76+ CONSUMER AGE GROUP USING THE POLARISCOPE

During the later stages of the main study it became important to know how elderly people perceived label copy and indexing arrows on child resistant caps and bottles in order to obtain information for the Innovative Child Resistant Packaging Systems grant study. A test of legibility was conducted using the polariscope only, with message II caps (nos. 2, 4, 6 and 8) and subjects in the 56-76+ age group. The subjects were residents of the Lansing area, all living independently in their own homes. The caps were presented in the same manner as in the main study. The individual results appear in Table 19, page 124 which contains raw data for the cap arrow, bottle arrow and the cap message. Note that in this age category, there were 31% who had visual acuity of 20/40.

Tables 20 and 21 show the results for the three age groups 18-25, 30-55 and 56-76+. A statistical analysis has not been done on the data to evaluate an age effect, but the regular decrease in legibility with age in each visual acuity level suggests the possibility of an age effect on the legibility measurements.

Table 22 contains the results when the age 56-76+ subjects observed indexing arrows on the caps and bottles of snap cap child resistant packages (see Illustration 11). This extension of the legibility test indicates that color contrast can be extremely helpful in improving legibility (or visibility) of important symbols. Note that for the

20/40 visual acuity, the color contrast provides a four fold improvement in legibility. Actually this improvement is greater than four fold because two of the four subjects with visual acuity of 20/40 could not see the bottle arrow at any polariscope reading (see Table 19). For these individuals, the maximum value of 90 degrees was assigned for computational purposes.

Table 19
RAW DATA: LEGIBILITY READINGS BY 56-76+ CONSUMER AGE GROUP USING THE POLARISCOPE (IN DEGREES OF ROTATION)

				Cap Arrow[1]				Bottle Arrow[2]				Cap Message[3]			
				W/W		B/W		W/W		B/W		W/W		B/W	
No.	Sex	Age Group	Visual Acuity	1	3	2	4	1	3	2	4	8	2	4	6
1	M	55-65	20/20	10	15	7	6	40	20	7	10	10	15	7	10
2	F	71-75	20/20	20	10	7	10	32	25	15	9	13	35	14	15
3	M	66-70	20/20	14	13	3	5	24	17	4	5	6	13	5	6
4	M	71-75	20/20	17	12	18	10	45	20	12	17	20	50	14	18
5	F	66-70	20/20	11	8	7	8	36	28	13	15	15	40	15	13
6	M	66-70	20/20	6	6	6	5	42	12	10	9	16	33	9	15
7	F	66-70	20/30	8	11	8	7	32	23	8	10	18	21	13	13
8	M	76+	20/30	25	19	20	13	25	25	26	12	25	55	19	28
9	F	55-65	20/30	12	8	4	5	*	20	9	7	50	*	17	45
10	M	76+	20/40	15	12	8	11	*	33	18	15	35	*	27	41
11	F	76+	20/40	13	13	12	10	25	20	12	14	49	*	29	35
12	F	76+	20/40	12	8	10	5	45	20	12	12	25	55	17	24
13	F	76+	20/40	13	17	11	12	*	25	16	15	25	*	19	24

1. Cap arrows nos. 1 and 2 are similar arrows with one in white-on-white (W/W), and the other in black-on-white (B/W). This is the same for cap arrow nos. 3 and 4. (See Illustration 11, page 127).

2. Bottle arrows nos. 1 and 2 are similar arrows with one in W/W and the other in B/W. This is the same for bottle arrow nos. 3 and 4. (See Illustration 11, page 127).

3. Cap message nos. 8 and 4 are two similar caps in 8 pts with one in W/W and the other in B/W. Nos. 2 and 6 are in 6 pts and are also in W/W and B/W. (See Illustration 2, page 29).

(*) The subject could not see or read the visual object at the maximum screen rotation. The maximum rotational value (90 degrees) was inserted in the calculation for the subjects.

Table 20

CAP TEXT LEGIBILITY (6 PTS) AT THREE LEVELS OF VISUAL ACUITY (56-76+ YEARS)

	20/20				20/30				20/40			
	W/W		B/W		W/W		B/W		W/W		B/W	
Age Group	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
18-25	11.0	1.1	7.0	1.6	15.0	5.3	9.0	2.0	-	-	-	-
30-55	16.0	1.8	9.0	2.5	32.0	16.2	20.0	10.4	-	-	-	-
56-76+	31.0	14.4	13.0	4.3	55.0	34.5	29.0	16.0	81.0	17.5	31.0	8.4

Table 21

CAP TEXT LEGIBILITY (8 PTS) AT THREE LEVELS OF VISUAL ACUITY (56-76+ YEARS)

	20/20				20/30				20/40			
	W/W		B/W		W/W		B/W		W/W		B/W	
Age Group	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
18-25	8.0	2.2	5.0	1.2	8.0	2.9	7.0	1.6	-	-	-	-
30-55	11.0	3.4	7.0	2.4	20.0	11.1	14.0	8.6	-	-	-	-
56-76+	13.0	4.9	11.0	4.2	31.0	16.8	16.0	3.1	34.0	11.4	23.0	5.9

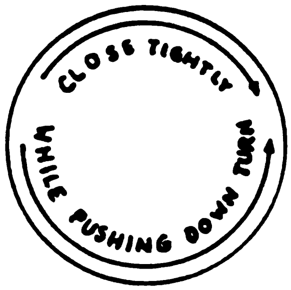
Table 22

CAP AND BOTTLE SYMBOL LEGIBILITY AT THREE LEVELS OF VISUAL ACUITY (56-76+ YEARS)

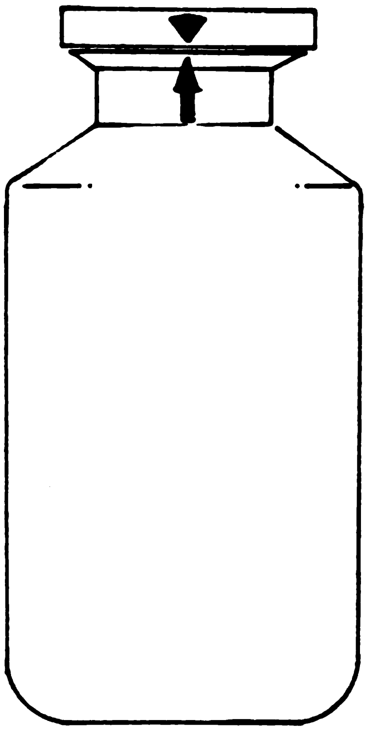
	20/20				20/30				20/40			
	W/W		B/W		W/W		B/W		W/W		B/W	
Age Group	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
CAP	13.0	5.1	8.0	5.1	15.0	8.9	11.0	8.3	13.0	1.3	10.0	1.7
ARROW	11.0	3.3	7.0	2.4	13.0	5.7	8.0	4.2	13.0	3.7	10.0	3.2
BOTTLE	37.0	7.7	10.0	4.1	49.0	35.7	14.0	10.1	63.0	32.8	15.0	3.1
ARROW	20.0	5.7	11.0	4.4	23.0	2.5	10.0	2.5	25.0	6.2	14.0	1.4

ILLUSTRATION NO. 11

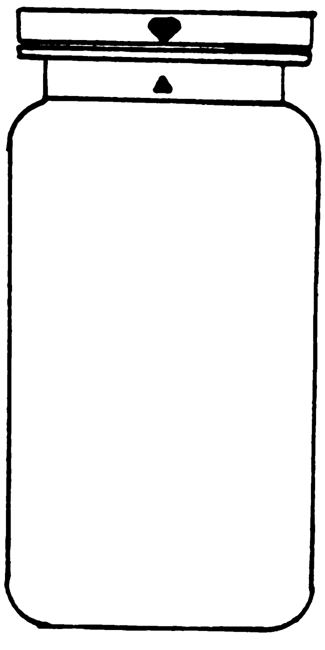
LEGIBILITY TEST SAMPLES



Closure text



Translucent bottle



White bottle