Analysis of variables measured in the visual screening of University of Birmingham freshmen and a critical review of visual screening methods and approach.

R. Targett

A dissertation presented for the degree of M.Sc. at the University of Aston in Birmingham

April 1970

SUMMARY

This investigation analysed the variation of the visual functions measured by the University of Birmingham visual screening technique, for the years 1958 to 1967.

A review of the history and development of screening techniques and survey of relevant literature was carried out.

The efficiency and repeatability of the University screening method was determined, firstly by questionnaire and re-screening follow-up methods, and secondly by a more controlled, experimental procedure. Comparison was made between this screening method and commercially available screeners.

In addition an evaluation of the sub-tests was made in respect of their accuracy, repeatability, sensitivity and specificity. Conclusions were reached as to which sub-tests should be included in the screening batteries, and which should be omitted because their contribution to the overall efficiency of the screeners was very limited.

A preliminary investigation was made into the effects on efficiency of changing the referral standards. The criteria for accepting any specific level of efficiency were discussed.

ACKNOWLEDGMENTS

I should like to thank the following people for their co-operation in this project:-
The British Optical Association for their financial support; the staff of the Health Centre and the Registry of the University of Birmingham; the staff of the Departments of Applied Psychology, Mathematics and Ophthalmic Optics of the University of Aston in Birmingham; in particular I would like to thank Professor G. V. Ball, Judith Hargroves and Michael Wolffe for their help and encouragement; Mrs. Bowles for her excellent typing, and all the students who volunteered their services as experimental subjects.

CONTENTS

1. INTRODUCTION

ORIGINS OF VISUAL SCREENING

The "struggle for existence" described by Charles Darwin in 1859 results in "natural selection" which produces a slow modification of species to suit their environment - "if any one species does not become modified and improved in a corresponding degree with its competitors, it will soon be exterminated". (Darwin 1859).

Similarly Man, in his struggle for existence, has produced his own selection techniques, which result in either the modification of the individual to suit his environment, or an attempt to change the environment to suit the individual. Selection techniques are used to classify the individuals and to classify the environments, so that they may be matched to produce the optimal combinations. Rao (1962) commenting generally on the choice of individuals for a particular task states:-

> "In many widely differing fields of activity schemes for choosing relatively small numbers of individuals from a large population, with the aim of finding those most suitable for some specified purpose, have been proposed and investigated. Common to all these situations is the need to distribute a large number of entities between two or more categories, on the basis of some trait incapable of exact measurement; all that can be done is to measure performance under standardised conditions, defined so that the measurement will be highly correlated with the ideal, and then to separate into categories in terms of this measurement".

Some examples of selection techniques which have been used can be found in crop selection in agriculture; methods of birth control; building systems; therapeutic value of different drugs in the treatment of various diseases; streaming of children in education; personnel placement in industry and the differentiation of non-diseased and diseased persons. These applications are concerned with the maintenance and development of a high food/population ratio, control of disease, and achieving the maximum potential from an individual by placing him in a situation which suits him.

One of the abilities by which individuals are classified is visual ability. Visual selection techniques have been developed to aid such classification. Their development can be attributed to those areas in which classification of individuals according to ability is necessary, in particular in industry, the armed forces, and education.

INDUSTRY

Ever-increasing industrial competition makes it desirable to find the most productive combination of person and environment, i.e. vocational selection according to the physical and intellectual attributes of the individual. Visual selection is an important part of this selection procedure. Its counterpart, concerning the environment in which the person may be placed, is visual task analysis.

Visual examination of the individual can be expected to increase the efficiency of the man-environment system in three ways:-

(i) By preventative medicine; the early diagnosis of disease may prevent absenteeism due to prolonged illness.

(ii) The correction by optical means of eye defects such as ametropia and heterophoria permits an estimation to be made of the worker's maximum visual capabilities Manpower potential may be wasted by assessing visual capabilities without first correcting ocular defects. This attitude is expressed by Jobe (1944):-

> "We are not concerned with the cause of the variation in visual performance, but only in the effect the variation will have upon industrial performance".

(iii) By assessing the visual capabilities of the individual he may be placed in a job to which he is suited. For example crane drivers should have good distance visual acuity and a high grade of binocular vision and stereopsis. Even more important than this is the avoidance of the situation in which a person is found doing a job for which he is entirely unsuitable e.g. an apprentice electronics engineer who is colour blind.

Following the establishment of a visual selection programme in industry one should expect to find people with healthy eyes performing tasks of which they are visually capable. The advantages in terms of increased productivity are as follows:-

i. Greater accuracy in performing tasks

If the operative can carry out his task with greater accuracy there is a reduction in the amount of faulty products which he produces. There will, therefore, be a reduction in the amount of waste caused by the rejection of these products. Accuracy also means that there are fewer accidents. Migliorino (1963) on examination of railway workers was able to show that groups of

workers of clearly diminished visual capacity gave rise to a considerably larger number of breakdowns than did those workers who enjoyed better visual capacity.

ii. Increased production speed

As a result of increased accuracy there will also be increased production speed, occurring both directly, by faster working, and indirectly, by cutting down the number of stoppages due to accidents.

iii. Reduced training waste

It is possible to predict which individuals are more likely to be successful at a particular task, and so only they should be selected for training.

iv. Reduced stress

There is less physical and mental stress placed on the individual. This is conducive to a happy working attitude.

These advantages of a visual selection programme are summed up in Fig. 1 : i.

THE ARMED SERVICES

International conflicts have made each nation strive to produce effective and efficient fighting forces. Here again selection procedures are used in the placing of personnel. Visual selection plays an important part in the classification of servicemen, and indeed determines whether they are accepted for service or not. Therefore the selection techniques are subjected to vigorous investigations so that their reliability is known. These investigations are usually well designed experiments with clearly defined aims, and the results obtained are useful contributions to knowledge in the field of visual selection. There are two main reasons for the success of

Visual selection programme in industry

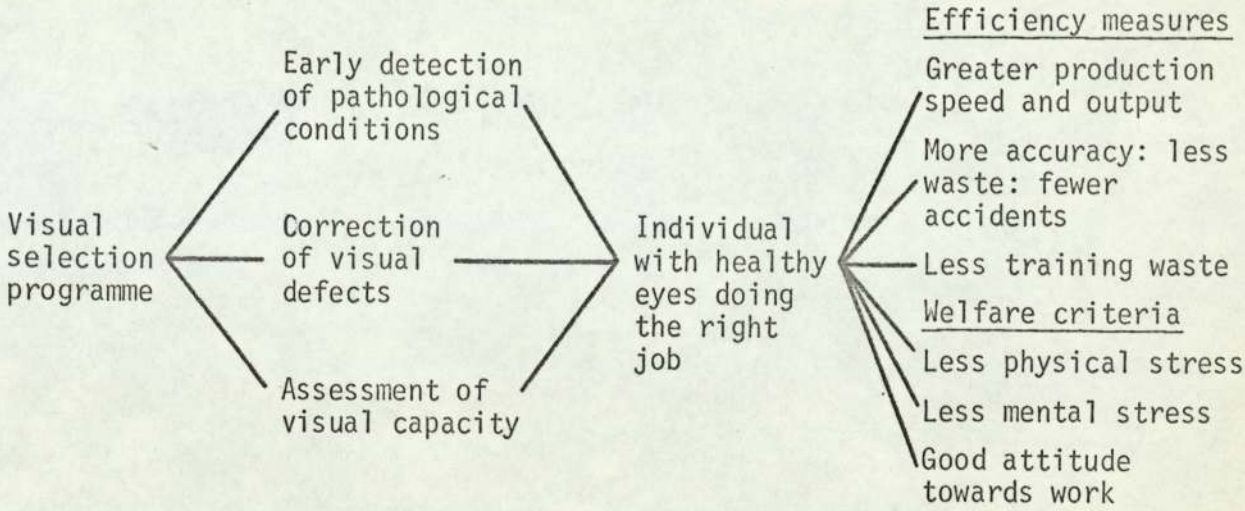


Fig. 1 : i

Visual selection programme in education

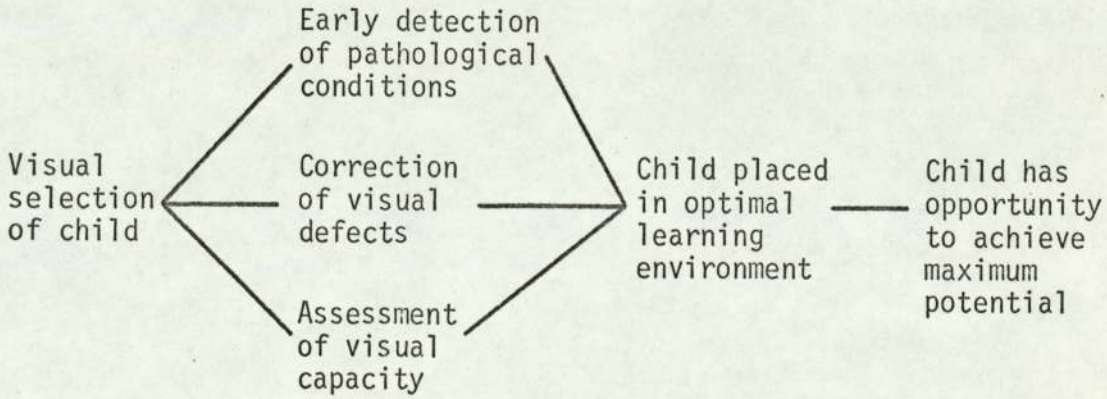


Fig 1 : ii

these investigations: firstly the experiments are designed by teams of investigators who pool their knowledge of various disciplines e.g. ophthalmologists, ophthalmic opticians, statisticians and psychologists: secondly, the investigations are carried out separately from the visual selection programme rather than as a subsidiary to it, this means that the selection procedure cannot put any constraints on the experimental design, and, therefore, the results are more reliable. This laboratory type of investigation introduces the uncertainty of making predictions about a real-life situation from its results. Chapanis (1967) advises caution in interpreting the results of such investigations. However a well executed study of this type must surely be regarded as preferable to a poorly controlled 'field' study.

EDUCATION

The concern for allowing an individual to develop his talents to the full has promoted the use of selection procedures early in his life, at school and before school, so that he may be channelled into the environment which will enable him to achieve maximum potential. The 1944 Education Act recommends that children should be classified according to their age, ability and aptitude.

Knowledge of a child's visual ability is important in classifying the child. The recommendations of the National Society for the Prevention of Blindness in 1961 stated that "If a pupil has poor vision, and this fact is not known, the difficulty affects his entire adjustment in school". A visual selection programme is essential in education. The advantages of this are summarised in Fig. 1 : ii.

SCREENING AS A METHOD OF SELECTION

The wisdom of making assessments of the visual abilities of both children and adults is now widely accepted. It is also accepted that ideally each individual should be given a full, clinical ophthalmic examination by a qualified person. This is stressed by Sloane and Savitz (1963) who state that "the ideal way to prevent and discover children's eye problems is to give every child a complete ophthalmological examination in the pre-school years and at regular intervals through the school years".

Unfortunately the cost of carrying out these examinations is prohibitively high, mainly because the majority of individuals examined do not require treatment and the cost of employing a professional person to make the assessments is very high. The "Plowden Report" (1967) commented upon the fact that thirty years ago doctors began to question whether a superficial inspection of large numbers of normally healthy children made the best use of their skill. The alternative is to employ a visual screening technique administered by a nurse or trained technician, with a professional follow-up examination for those individuals found to require it.

DEFINITION OF VISUAL SCREENING

Sloane and Savitz in 1963 gave a definition of screening in general, which can equally well be applied to the subject of visual screening. They state that:-

> "A screening test is a relatively simple, short and inexpensive test which can be administered reliably by non-professional testers to a large population in order to detect those members of that population who may require professional diagnosis and care".

2. REVIEW OF LITERATURE AND MAJOR DEVELOPMENTS

In the seventeenth century Bernadino Ramazzini made a study of health in relation to occupation. He was the first to show concern for the eye health and sight of persons in industry, and the concept of industrial medicine can be attributed to him.

In education the individual was at first entirely responsible for his own medical welfare; official concern for student health can be attributed to Amherst College (U.S.A.) in 1856, who made an official statement that "proper measures" could be devised to avoid much "breaking down of the health of students". The measures which were taken at first were lectures on health education combined with extensive physical education programmes, later full health services were set up in colleges. For example, in 1902 Harvard University erected a large infirmary as part of it's student health programme.

Snellen's "Optotypes" introduced in 1862 provide a repeatable method of recording and estimating visual acuity, thus providing a basic sight screening test. There is no record of it being used as such until Connecticut initiated a state supported school vision testing programme in 1899.

The Landolt Ring was established as a standard test chart at the International Ophthalmological meeting in 1909. It was this sort of concern for standardisation of testing procedures which made vision screening possible.

The 1918 Education Act (Fisher) set up medical inspection teams, these teams carried out the first mass inspections of vision in this country.

In 1934 Betts devised the first vision screening device. It was in the form of a Brewster stereoscope with a series of slides for the examination of visual acuity, fusion, stereopsis and other visual functions. Originally it was designed for screening school children as the Betts Ready to Read Tests, but a series of special stereograms has since been developed for use in industry, the present form of this instrument is the Keystone Telebinocular. Because more than just visual acuity was examined Betts tests were considered by Sloane and Rosenthal (1960) as the first major advance in visual screening since the introduction of the Snellen chart.

It was found, however, that Betts Tests referred 85% of the children tested. Peters (1938) believed that a large proportion of these were over-referrals. . This provoked antagonism towards the tests from the ophthalmic profession, the teaching profession and the children's parents. Lancaster (1939) initiated an evaluation of the Betts tests. As a result of this study the American Medical Association pronounced that Betts tests were unacceptable for screening children.

In Great Britain the Factories Act of 1937 stated that juveniles (14 to 16 years of age) "should be examined by the factory surgeon and an eyesight test be carried out". This established the need for an efficient visual screening test in industry.

The Snellen test was widely used as a screening test. Spache (1939) was one of many who questioned the validity of using only a visual acuity test. A study of the Snellen test as a screening procedure was carried out by Peters in 1961. He concluded that the "Snellen Test is not an adequate nor efficient method of finding those children in an elementary

school population who have vision problems".

Sloane in 1940 developed the Massachusetts Vision Test (hereafter abbreviated to M.V.T) for school children. It was a subsidiary development arising from a study into the causes of early failure at school. The test included sub-tests to detect errors of binocular vision and latent hypermetropia, as well as reduced vision. The basic screening procedure has since gained wide acceptance and in 1947 was approved by the American Medical Association as a suitable method of screening school children.

The sudden need for skilled labour in 1940 due to the demands of war stimulated interest in visual screening tests which could be used to ease the burden of job placement. The Orthorater was developed between 1940 and 1943 by Tiffin, Wirt, Kuhn and Shephard in conjunction with Bausch and Lomb. At about the same time the American Optical Company Sight Screener made its appearance, and both instruments were leased to various commercial and Government organisations. These instruments were not readily accepted by clinicians. At the American conference on industrial ophthalmology (1945) Berens and Presti said that "no ideal optical instrument is at present available for rapid eyesight examinations in industry ..... Until a reliable vision screening instrument is available the Snellen test type, Maddox rod and Maddox wing and the other orthodox methods are still recommended as the most suitable for eyesight examinations of employees in industry".

The original versions of the sight screener and Orthorator were considered for use in schools. Studies of their reliability

in the school situation have been made, notably by Crane, Foote, Scobee and Green (1954), they found them to be unreliable because of their complexity.

Bolton (1951) reported upon the mass medical screening established in 1949 at the University of Birmingham. This included a vision screening section. This consisted of a distance visual acuity and phoria test using a Turville Infinity Balance unit; a test for hypermetropia; the Turville Near Balance unit; a near phoria measurement and an Ishihara colour vision test. The assessment of the subject is made by an ophthalmic optician who bases his conclusions not only on the results of the tests, but also on the subject's reported history and symptoms.

Diskan developed the Atlantic City Eye Test in 1952, as an "assault against inadequate visual screening programmes in schools". This is basically the M.V.T. with modified phoria tests and referral standards. Diskan (1955) claimed that it has a lower over-referral rate than the M.V.T.

One of the main difficulties encountered in visual screening is to find the optimal standards for referral. In an attempt to find these ideal standards for school children, a committee headed by Lancaster (1954) was set up. Questionnaires were sent to 149 ophthalmologists and information was obtained regarding the standards which they used in their own practices and the standards which they would advocate for use in a screening battery. The committee concluded as a result of this survey that among American ophthalmologists considerable differences of opinion exist in regard to the standards which they would set for the referral of school children to them. The committee recommended that further study of the problem was desirable.

The California State Recommended Procedure of 1953 has the basic three parts of the M.V.T. but a cover test is used for testing muscle balance, rather than the Maddox rod.

The St. Louis study reported in 1954 that the M.V.T. was the least inefficient of a number of screening procedures used in schools, as a result of this report the companies marketing screening instruments brought out modifications of the batteries of tests already in production, to comply with the M.V.T. tests and standards, for use in schools. In 1955 the American Optical Company released the Massachusetts School Vision Test modification of the original sight screener; in the same year the Telebinocular was modified to the standards of the M.V.T.; and Bausch and Lomb introduced the Modified School Orthorater in 1956.

The next major development in vision screening was made in 1959, when the Modified Clinical Technique was developed (hereafter referred to as the M.C.T.) in the Orinda Union Elementary School District. This M.C.T. resulted from a three year research project consisting of a longitudinal study of the visual status of approximately 1,000 children. It is performed by a professional examiner who obtains an estimation of refractive error by retinoscopy; measures muscle balance; and examines the eye for disease both externally and internally. The Orinda Study (Blum, Peters and Bettman 1959) concluded that:-

> "By far the most effective screening method was the Modified Clinical Technique. It was the only screening method that discovered essentially all the children with visual abnormalities who needed professional attention, and with a minimum of needless referrals. It uses a few tests that

cover a wide range of problems. Both the reliability and the validity of the tests are excellent".

In producing an effective screening technique the premise of using non-professional testers is abandoned. The wisdom of doing so has produced a point of controversy.

The Vision Tester was introduced by the Titmus Optical Company in 1959, to be used for testing adults in industry. More recently the Rodatest (Rodenstock) and the Rapid Vision Tester (Carl Zeiss Jena 1964) have joined the ranks of screening instruments for industry. These instruments test the same basic visual skills, but use slightly different methods of measurement.

It is apparent that most of the development work on visual screening has been carried out in the United States.

Apart from the major developments described above, there have been many studies of the various techniques in use. This wealth of literature may be divided into four main types of study:-

(i) Reports on screening programmes operating in specific situations. These usually state how many individuals were screened, how many failed, and how the failure rate compares with the same screening test in a different population e.g. children of different age groups, or a different test on the same population. These reports are useful mainly in lay publications, in order to educate the public, or a specific section of the public, for example teachers and parents. The reason why there has been greater development in screening in the United States than in this country is probably because

there is more public awareness of the need for visual screening, due to the profusion of this type of report. In this country probably the most publicised studies of this type were the ones made by Cutler and Davey (1965) and Unger (1962) on the vision of drivers.

Some more examples of this type of investigation are:- Macrae 1956: Michigan Department of Health 1965: Martin 1964: Cinotti and Siegel 1964: Kushner 1953: Leverett 1957: Sloane and Gallager 1950.

(ii) Studies where an attempt is made to find the efficiency of a screening technique by following up the referred cases and finding the proportion of incorrect referrals made by the screening. They may also make comparisons between several screening techniques. These investigations show only half the facts, as a true estimate of efficiency involves following up the subjects who pass the screening as well as those who fail.

Some examples of this literature are:- Murphy and Thyng 1957: Diskan 1955: Blackhurst and Radke 1964: Reese 1964: Gutman 1956: Sloane and Gallager 1952: Yasuna and Green 1952.

(iii) Investigations where the whole population (or a random sample taken from the population) are given a full clinical examination by a professional person, as well as one or more screening tests, so that an estimation of the proportion of under-referrals as well as over-referrals can be made for each technique under study. This type of study produces a useful estimate of the efficiency of visual screening techniques but it is necessarily longer and more expensive than the previous two types of study, and involves the co-operation of more people.

Examples of this most useful type of investigation are:- Imus 1950: Agarwal and Das 1964: Morgan, Crawford, Pashby and Gaby 1951: Cox 1967: Gentile and Johnstone 1961: Roberts 1963: Crane, Scobee, Foote and Green 1952: Robinson 1953: Gordon, Zeidner, Zagorski and Uhlaner 1954.

(iv) Discussions of hypothetical screening situations, and the theory underlying screening procedures. These discussions are important in that they demonstrate the limitations of screening. They also suggest ways in which the tests can be modelled to suit a particular screening situation.

Examples of this literature are:- Rosenbloom 1955: Neyman 1947: Vecchio 1966: Finney 1962: Thorner and Remein 1961: Sproul 1966: Van Woerkom and Brodman 1961.

The conclusions reached when reading these reports are the same as that found by the St. Louis Study (Crane et al 1952):-

> "The findings of the study ... do not permit a conclusion that any one procedure is superior to the others. The screening programme that is best in one situation may be less suitable for another".

This was exemplified by Murphy and Thyng (1957) who in discussing the usefulness of employing mechanical means for screening school children stated:-

> "The characteristic which supposedly makes them superior to other methods of testing is the very thing which makes them unsuitable for use with children - their comprehensiveness and complexity".

Hence it can be seen that there is not one overall efficient method of screening which can be used in any situation. Rather there are many screening methods, each being suitable for a specific type of situation. The choice of screening

method is determined mainly by the type of population to be screened and the resources available.

To determine whether the screening method chosen is efficient certain criteria may be applied to it. It is in discussing these criteria and their application that the previous studies are particularly helpful.

3. BACKGROUND TO PRESENT STUDY

The intellectual and material growth of the society in which we live, depends upon it achieving a reasonable return from the investment of its resources. When these resources are invested in individuals, the society must take every precaution to ensure that the individual is a "good risk". For example, money and specialised teaching invested in a young person at University expects in return that the undergraduate will complete the course in the time specified, and having done so will use the knowledge and experience gained on the course for the benefit of society. Failure to do this results in considerable cost both to society and to the individual, particularly if failure occurs at a late stage in the course.

The failure may be due to an obstacle which could have been foreseen or avoided, and it is up to the society to discover and eliminate as many of the obstacles as possible so as to minimise waste. The University of Birmingham has tackled this problem by medically screening it's students early in their course. This does not reveal all the potential "drop-outs", as reasons for a student not completing the course are not always medical. They may, for example, be psychological or social.

Screening started at the University of Birmingham in 1945, when students were given individual, forty minute medical examinations, with technical tests. The work was done by the University medical officer and a technical assistant, over a period of six months.

In 1949 the system was changed, mainly because of administrative reasons. The new system was described by the University Medical

Officer (Bolton 1952) as a station-type overhaul, designed on "conveyor-belt" lines with students passing at the rate of one every two minutes, the whole intake being surveyed in a single week, by a large and mostly semi-skilled team. The "stations" cover a variety of medical investigations including: height; weight; mass radiography; dental checks; tests for anaemia; and visual tests.

The present study is concerned with the small number of "stations" which together form the visual screening tests. The tests are based on Sloan's Massachussetts Vision Test (1940) modified to utilise the Turville Infinity and Near balance units. These tests have been used, with slight modification, for the past twenty years, and this study will attempt to evaluate them.

The two main objectives of this investigation are:-

(i) To analyse the results of the visual screening of University of Birmingham "freshmen" over the years 1957 to 1967.

(ii) To find out whether the screening programme used was the most suitable for the situation, and if not, to make modifications based on the results of these investigations.

4. UNIVERSITY OF BIRMINGHAM VISUAL SCREENING

There are nine small sections forming the complete visual screening tests. If, when a student is screened, the results of the tests fall within the specified limits, then he is assured that all is well. However, if the results are outside these limits, the student is referred either to the University Health Service or to the Optician of his choice for a thorough eye examination.

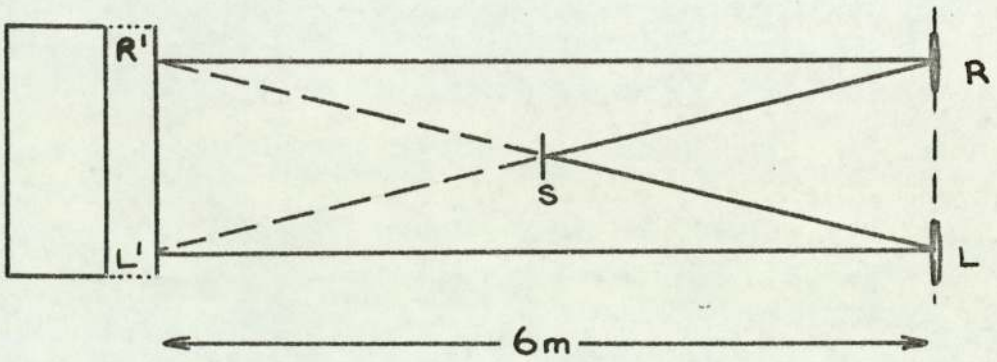
The screening consists of the following:-

1. The student is asked whether he normally wears spectacles or contact lenses. If he does they must be worn for the tests.
2. He is asked when his last eye examination took place: (a) during the last year (b) previously or (c) never. School and military service tests are disregarded for this purpose.
3. A simple Turville Infinity Balance unit with 6/12 and 6/7.5 letters is viewed through +2.00 D lenses to determine the presence of hypermetropia. The T.I.B. unit allows the eyes to be dissociated yet allowing simultaneous testing of each eye (see fig. 4 i).

If the student is +2.00 D or more hypermetropic he will be able to see the letters clearly, as the lenses correct the error, and relax his accommodation. If he is myopic, astigmatic or emmetropic the +2.00 D lenses will blur the letters so as to make them unrecognisable. To pass this test no letters should be read, or the 6/12 letters read equally well with each eye.

4. The previous test is repeated without the presence of the +2.00 D lenses, so that the visual acuity may be ascertained. All letters must be read correctly in order to pass the test.

SCHEMATIC DIAGRAM OF THE TURVILLE INFINITY BALANCE UNIT



The system is viewed from the right. The septum (S) is placed so that the right eye (R) only sees the right side of the chart (R') and the left eye (L) only sees the left half of the chart (L'). In practice the septum is placed on a mirror, and the letters reversed on the chart to be viewed through the mirror, and placed in the same plane as the subject's eyes. This shortens the distance required to set up the test.

Fig. 4 i

5. If only one set of letters is read correctly, or the letters are jumbled and confused, denoting an exophoria just large enough to superimpose the image from each eye, then a 3 prism base in is applied to one eye to compensate for the effect of a small exophoria, and separate the sets of letters. This is merely an extension of the visual acuity test. If both sets of letters can be read correctly the student passes the visual acuity test, but a cover test must be undertaken later in the series of tests (test 9).

6. Near test. A Turville Near Balance unit is used, in which the two eyes are dissociated to provide a check on near visual acuity for each eye simultaneously. All the words must be read correctly in order to pass the test.

7. Near lateral heterophoria as measured by a convergiometer. In a convergiometer the eyes are dissociated by vertical prisms, the subject sees two images of a scale with a central vertical arrow. He must state which letter on the scale the lower arrow points up to. If a reading of more than 10 Δ exophoria or 4$^{\Delta}$ esophoria is recorded, the cover test must be carried out in test 9.

8. A set of four Ishihara plates is used to screen for abnormal colour vision. Failure of this test does not constitute a failure of the entire screening, but the student is warned of any difficulty he may experience on his course due to defective colour vision.

9. The last station of the series is an assessment of the student by an ophthalmic optician. He reviews the results of the screening together with any further examinations which the results suggest e.g. cover test, ophthalmoscopy etc. The

results taken with history and symptoms enables him to assess the student's visual state in general terms utilising a scale graded as follows: pass; satisfactory; referred to own optician; referred to the University Optician next term; referred to the University Optician this term; immediate referral either to the University Optician or Doctor, or to hospital.

The aim of this procedure is speed, but with minimum loss of accuracy. A subject should pass through these nine "stations" in about nine minutes, i.e. one per minute.

5. ANALYSIS OF VISUAL SCREENING TEST RESULTS AT THE UNIVERSITY OF BIRMINGHAM. YEARS 1958 - 1967

The screening results for each student are kept on his medical record at the University Health Centre. From analysing these results it was possible to find the incidence and variation of visual characteristics over a number of years.

The screening records for the years 1958 - 1967 were coded and put on to punched cards to facilitate sorting. A total of 10,936 student records were coded. The code used was that described in Appendix II. The cards were sorted to find yearly variations.

The following graphs and tables show the variations in percentages:

Figs. 5i - 5x.

χ^2 tests applied to all the variables distributed over the ten years showed that they had all varied significantly in that time. The p-values found were all smaller than 0.001.

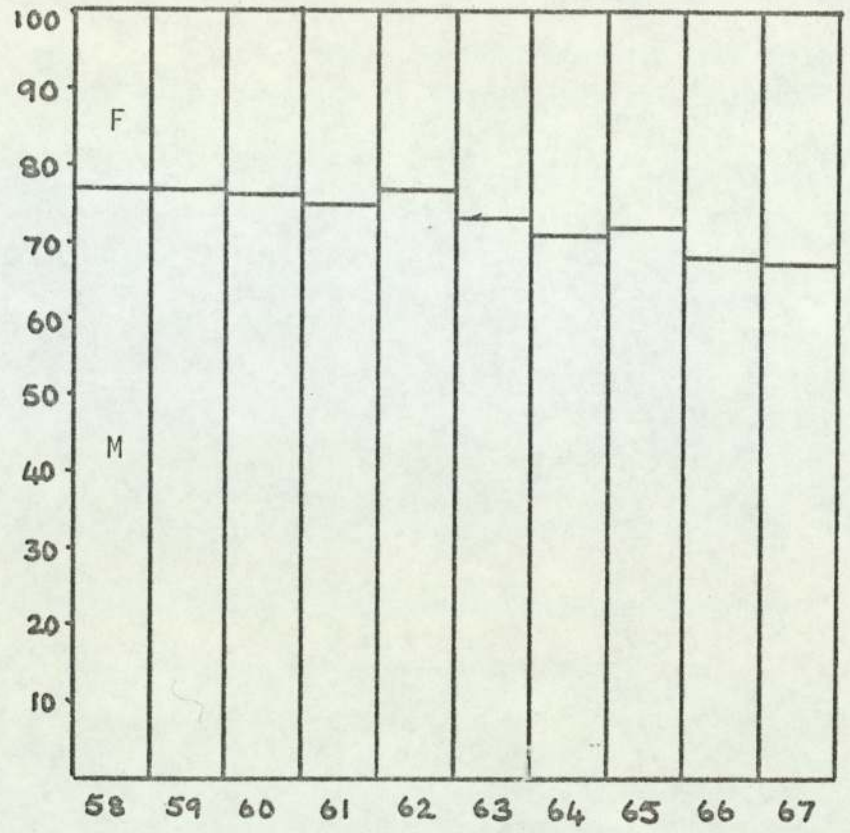


Fig. 5.i. Sex of students

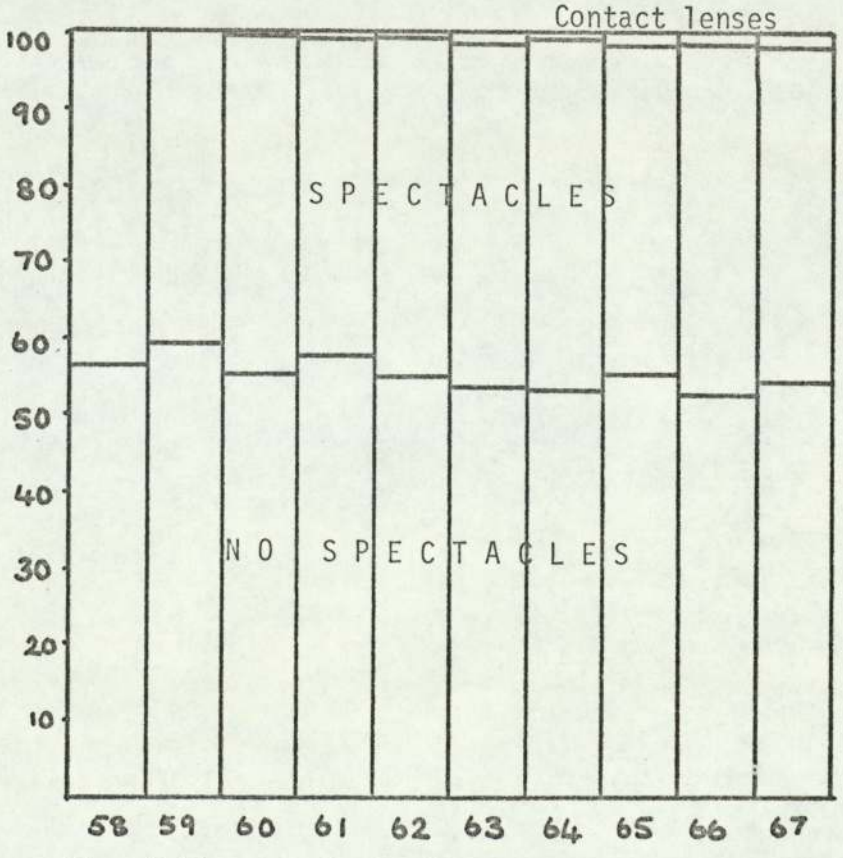


Fig. 5.ii. Corrections worn

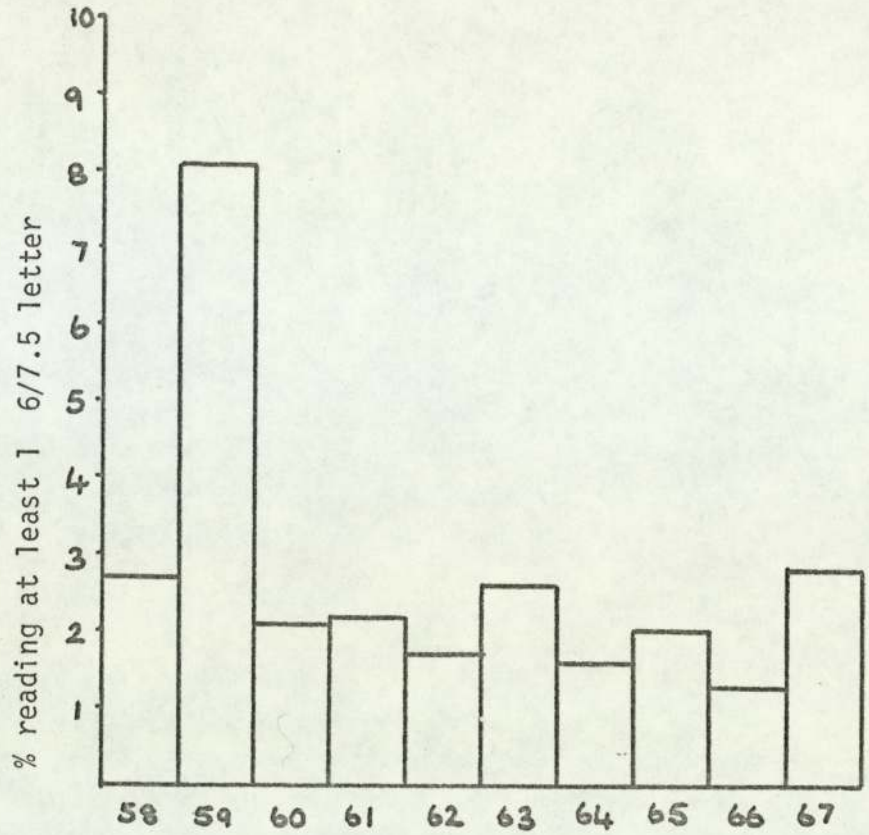


Fig. 5. iii. Presence of uncorrected hypermetropia L.E.

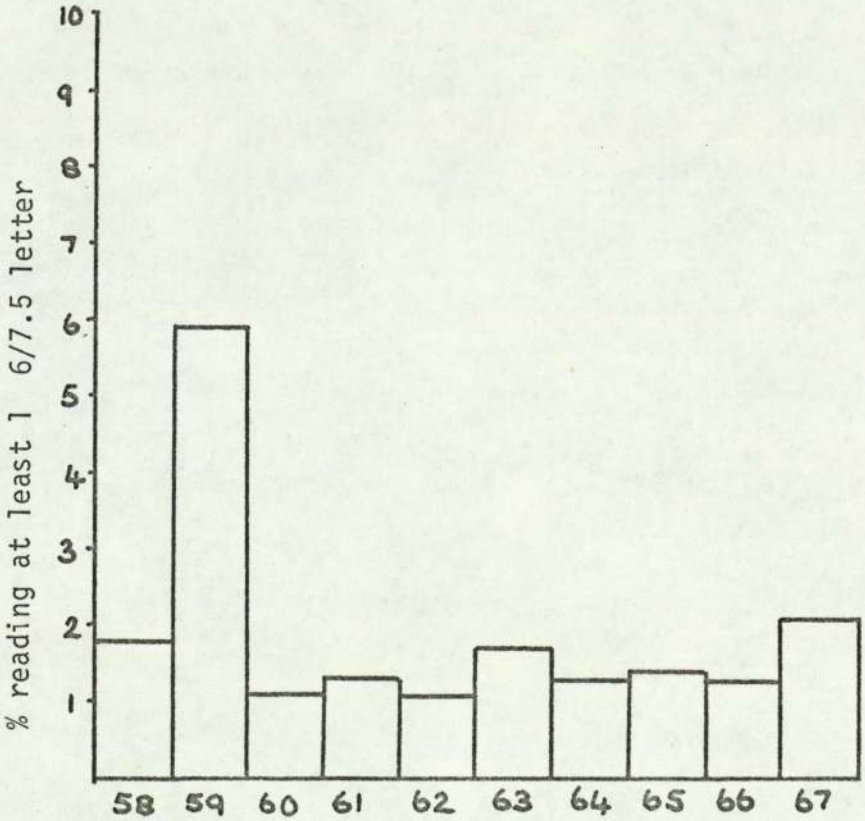


Fig. 5. iv. Presence of uncorrected hypermetropia. R.E.

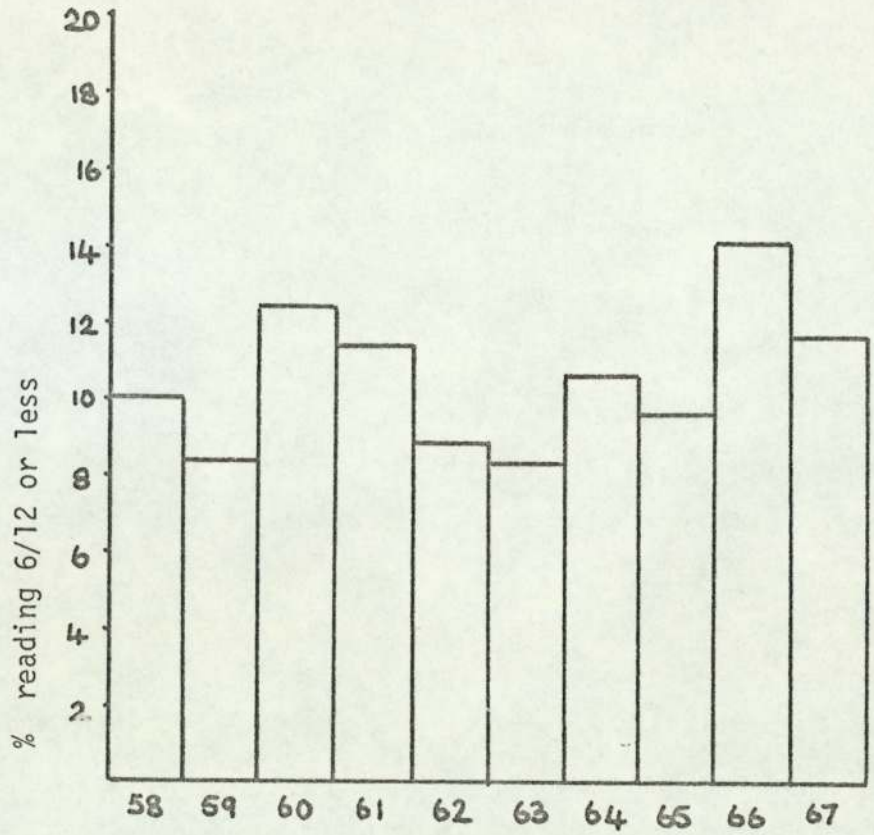


Fig. 5.v. V.A. L.E.

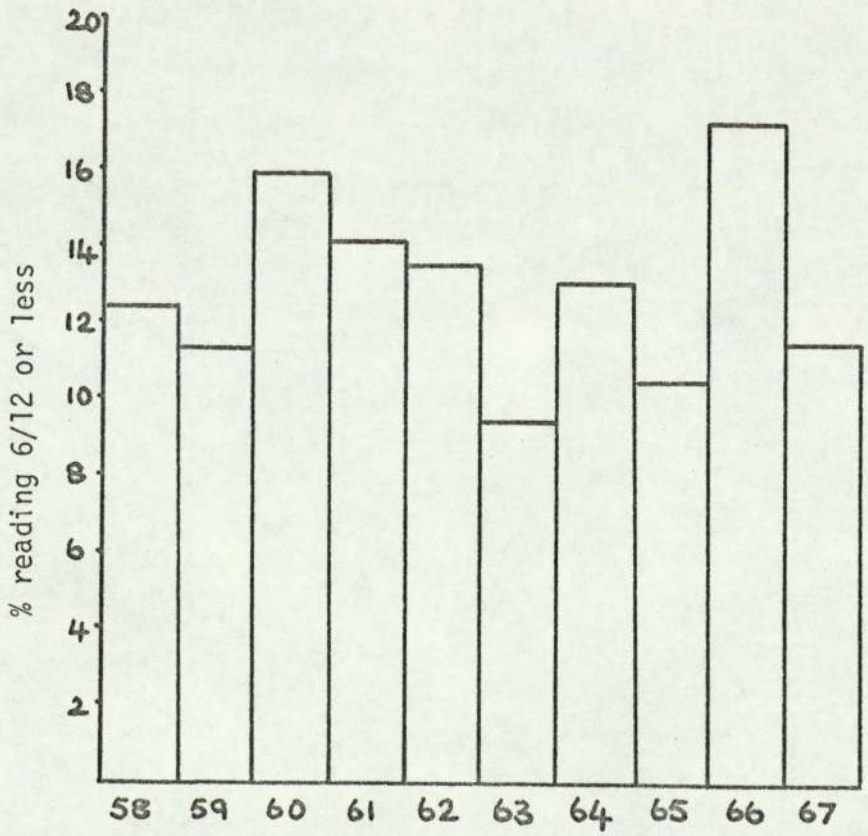


Fig. 5.vi. V.A. R.E.

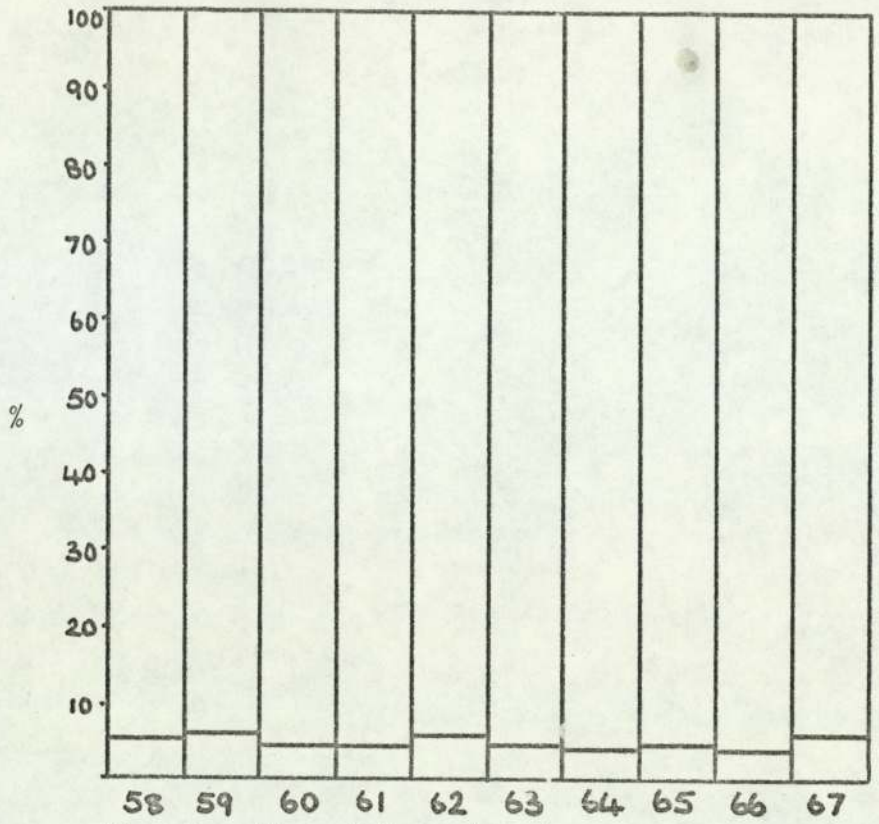


Fig. 5. vii. Colour deficiency

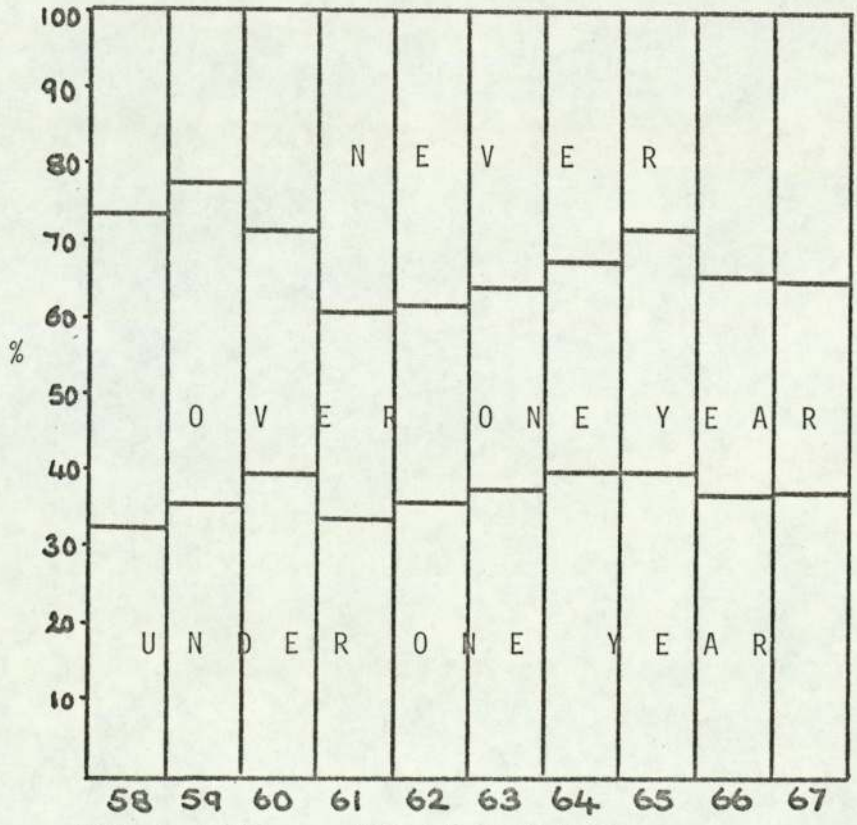


Fig. 5. viii. Date of last test

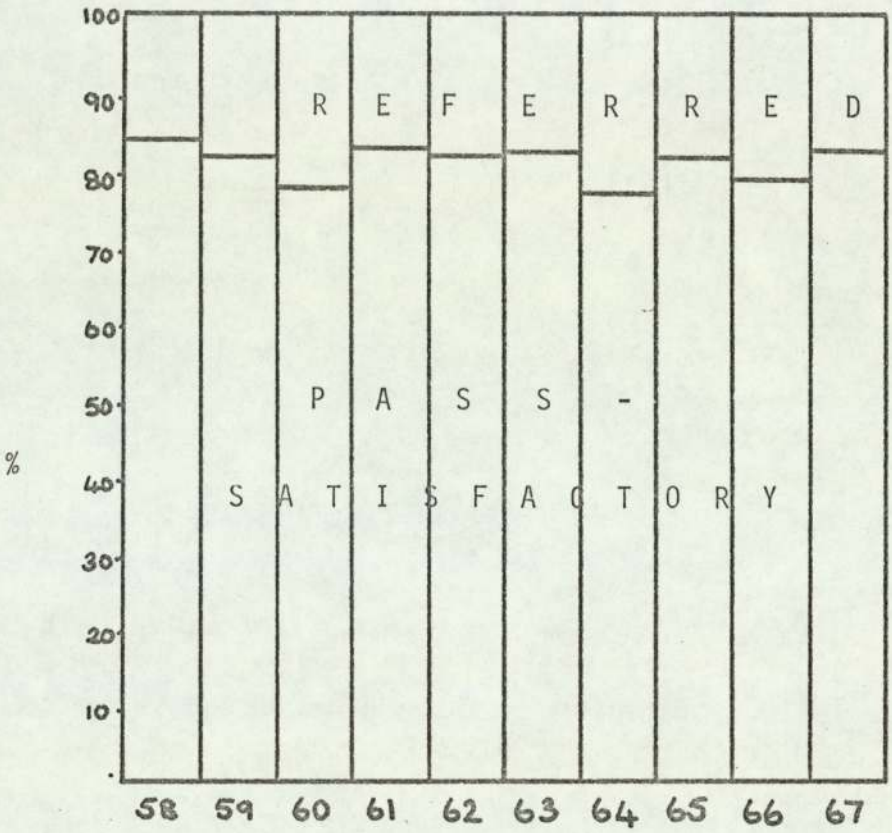


Fig. 5. ix. Assessment

YEAR

CONVERGIOMETER READING

	58	59	60	61	62	63	64	65	66	67
I	16.6	2.5	2.3	3.6	2.6	2.3	3.2	3.3	2.6	3.7
R	0.1	0.2	0	0.2	0.2	0.3	0.2	0.2	0.4	0.6
S	0.1	0.6	0	0.3	0.2	0.4	0.2	0.4	0.1	0.2
T	0.9	0.2	0.3	0.5	0.2	0.8	0.7	0.7	0.9	0.6
U	0.9	1.1	0.9	1.2	1.7	1.4	1.9	1.2	1.4	1.3
V	1.3	2.2	1.8	2.0	1.6	2.4	1.6	1.5	2.4	2.0
W	4.7	4.0	3.6	3.9	3.9	4.2	3.8	3.3	4.9	3.5
X	6.9	8.4	6.9	9.2	7.0	7.2	7.3	7.1	7.6	7.2
Y	12.5	14.7	13.5	12.7	12.8	15.6	10.2	13.6	12.6	13.6
Z	25.0	29.1	24.7	28.4	24.7	27.4	25.7	25.3	24.0	29.0
O	30.2	23.3	32.0	27.8	30.5	23.1	30.9	24.9	25.0	19.3
A	9.1	9.4	9.0	7.0	10.6	9.7	8.8	12.0	12.1	13.0
B	2.6	1.7	2.7	1.8	2.0	1.9	2.2	2.8	3.1	3.3
C	2.1	0.6	1.5	0.7	1.1	1.0	1.6	1.5	1.4	1.6
D	0.7	0.6	0.6	0.3	0.7	0.7	0.8	1.3	0.9	0.7

Fig. 5.x. Percentage distribution of convergiometer readings

6. APPRAISAL OF THE UNIVERSITY OF BIRMINGHAM VISUAL SCREENING METHOD

In order to make an assessment of a screening technique two questions must initially be asked:-

1. How efficient is the technique?
2. How precise are the measurements it makes?

The following is a discussion of both these questions.

1. Efficiency

It has been said (Murphy and Thyng 1957) that a perfect screening test is one which finds all people needing further care and refers no-one unnecessarily. Unfortunately the perfect screening test is impossible to achieve because, as Lippmann (1962) has stated, screening methods are basically procedures designed to be applied by a semi-skilled person, not an expert, to sort out those people who probably have abnormalities, from those who probably do not. They are carried out without the exercise of judgment by a qualified expert. Professional knowledge is, therefore, substituted by such tests. It is this substitution which accounts for the errors in visual screening. The errors can be attributed to four main factors:

i. Because the screening unit must have a limited number of sub-tests, the number of visual characteristics tested is restricted. The sub-tests chosen for inclusion cannot contain tests for all defects; to do so would make the screening procedure too long and complex. Those defects which are not looked for must inevitably be missed.

ii. The tests themselves may be inaccurate.

iii. The referral standards, that is the pass/fail cut-off points, may not be at optimal positions.

iv. The subject's responses may be misinterpreted by the person applying the screening test. Also, decisions made on borderline cases may be erroneous.

These sources of error are inherent in any true screening test. Because of these errors a population will, after being screened, always be divided into the following four groups:-

A. TRUE REFERRALS

Those people referred by the screening who do require professional care.

B. OVER-REFERRALS

Those people referred by the screening who do not require professional care.

C. UNDER-REFERRALS

Those people passed by the screening who do require professional care.

D. TRUE PASSES

Those people passed by the screening, who do not require professional care.

The only criteria we have for assessing the efficiency of a screening technique are the proportions of these four groups. To find these proportions follow-up examinations must be carried out on a random sample of subjects taken from the population.

Crane (1950) pointed out that in assessing the efficiency of a screening programme one must compare its results with a standard. This standard is the decision of the specialist,

ophthalmologist or optician, as to whether or not the subjects selected do need observation or treatment because of any ocular condition. One method of comparing the results is to find the correlation between the clinical results and the screening results. The phi correlation coefficient is most useful for comparing different screening techniques, but is not very informative about the technique itself.

The phi coefficient of correlation (Φ) as a measure of screening efficiency

The phi coefficient, Φ, is used most often to measure association between variables which are expressed dichotomously, e.g. alive or dead, pass or fail. It makes no assumption regarding the normality of the distributions of the variables. The numbers of the four groups of people, A : B : C : D, can be placed on the following matrix to facilitate computation of the Φ coefficient.

	Required professional care	Does not require professional care	
Referred by screening	A	B	(A + B)
Not referred by screening	C	D	(C + D)
	(A + C)	(B + D)	N

Table 6.i.

$$\text{The phi coefficient of correlation } \Phi = \frac{AD - BC}{\sqrt{(A + B)(C + D)(A + C)(B + D)}}$$

If there is perfect agreement between the assessment made by the screening and by clinical methods, then $\Phi = 1$; if $\Phi = 0$ there is no such agreement.

This measure is used by Cox (1967), the Orinda study, and the St. Louis study. This is, therefore, the measure that has been used here.

Sensitivity and specificity of a screening test

The phi coefficient does not provide much information about the screening technique itself. Using the same four groups, A : B : C : D, we can extract more useful information about the screening test by finding its sensitivity and specificity.

Thorner and Remein (1961) define sensitivity as "the ability of a test to give a positive finding when the person tested truly has the disease under study". Expressed in terms of the four groups, as a percentage:-

$$\text{Sensitivity} = \frac{\text{True referrals}}{\text{All subjects who require care}} \times 100\%$$

$$= \frac{A}{A + C} \times 100\%$$

Collen et al (1964) called this the "diagnostic ability" of a test to detect the presence of an ABNORMALITY.

Specificity is defined by Thorner and Remein (1961) as "the ability of the test to give a negative finding when the person tested is free of the disease under study". Expressed as a percentage:-

$$\text{Specificity} = \frac{\text{True passes}}{\text{All subjects who do not require care}} \times 100\%$$

$$= \frac{D}{B + D} \times 100\%$$

This was called the "power of the test" to detect the presence of NORMALITY, by Collen et all.

Specificity and sensitivity may best be demonstrated graphically, using the theory of overlapping distributions. According to this theory the population being screened in reality consists of two groups, those requiring care and those not requiring care: both groups possessing the attribute being measured by the screening, with different frequencies at various test values, see fig. 6.i. Persons whose screening value falls between points X and Y cannot be assigned definitely to one group or another on the basis of this test result alone.

For example, when screening for glaucoma using intra-ocular pressure as the attribute to be measured, the two overlapping distributions are for glaucomatous eyes and non-glaucomatous eyes, see fig. 6.ii. An eye whose intra-ocular pressure falls within the values X and Y can be assigned either to the distribution of glaucomatous eyes or to the distribution of non-glaucomatous eyes. Therefore, there is no cut-off point for which it can be said "eyes with an intra-ocular pressure above this value are glaucomatous, those below this value are not glaucomatous". However, screening tests must have a cut-off point so that a person may pass or fail. If the cut-off is placed at point O, then most of the glaucomatous eyes will be failed (group A, true referrals), and also some non-glaucomatous will fail (group B, over-referrals). Similarly some glaucomatous eyes will be passed by the screening (under-referrals, group C) and most of the non-glaucomatous eyes will be passed (group D, true passes). As sensitivity is the proportion of group A to groups (A + C) then it is shown by the proportions under the distribution

curve for glaucomatous eyes, fig 6.iii. Similarly, specificity is the ratio between group D and groups (D + B). This is the proportions shown under the distribution curve for non-glaucomatous eyes, see fig.6.iv.

For quantitative tests such as the measurement of intra-ocular pressure, it is possible to change the sensitivity and specificity within the established pattern by changing the cut-off point of the screening test. However, an increase in one value will decrease the other. For example, if the cut-off point is placed at point X, then the ratio $\frac{A}{A + C}$ will become large, i.e. increased sensitivity, but the value of $\frac{D}{D + B}$ will be reduced, i.e. reduced specificity. The opposite effect is produced by placing the cut-off point at Y.

If the prevalence of the abnormality is changed so as to alter the shape of the distribution curves, then the sensitivity-specificity pattern will also be changed, even though the test and the cut-off point remain the same. Therefore, care must be exercised when applying the same test to different populations, in case the prevalence is not the same within the populations.

Evaluation of the efficiency of the University of Birmingham screening technique

The efficiency of the screening technique used at the University of Birmingham, can be evaluated by placing students in the four categories, true referrals, over-referrals, under-referrals and true passes. This requires that the results of a clinical examination of their eyes is known.

It was at first considered useful to study the ophthalmic records of students who had been referred by the screening and

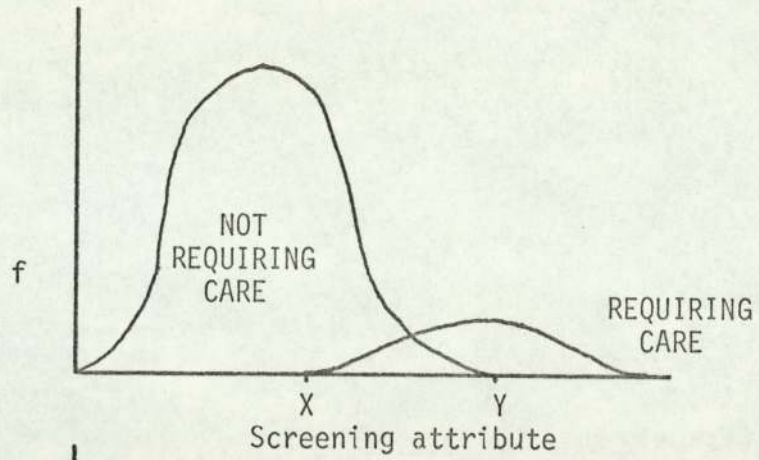


Fig. 6.i.

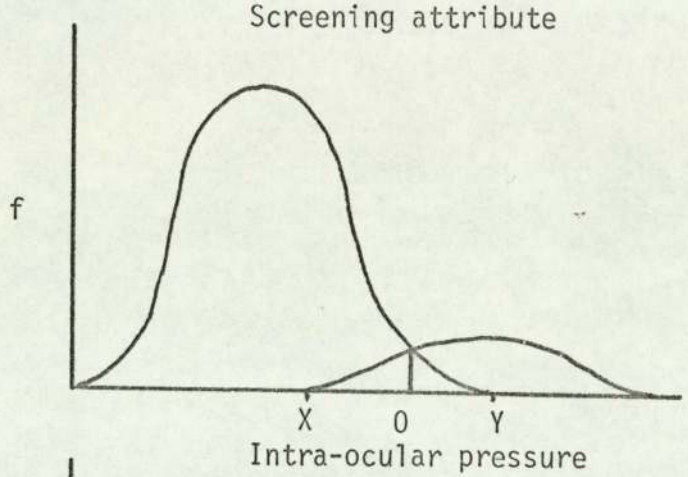
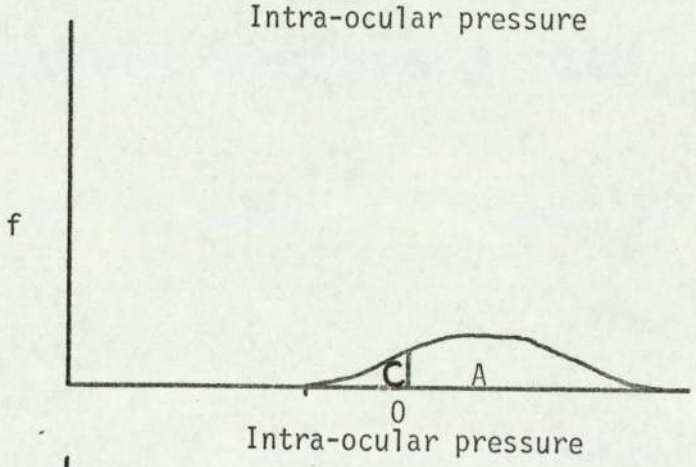


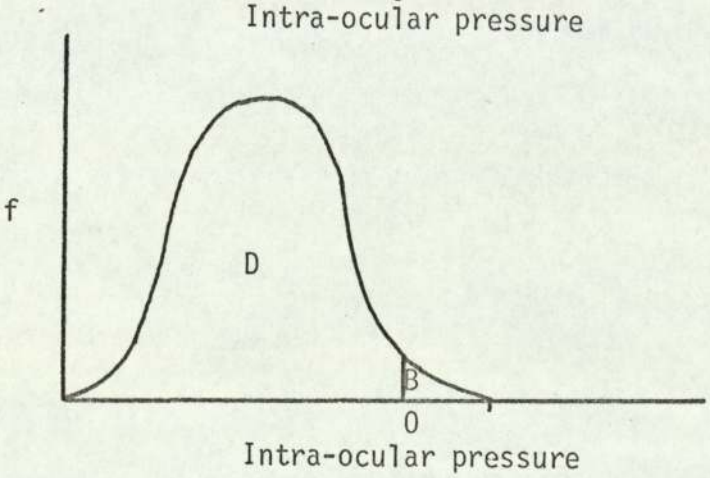
Fig. 6.ii.



SENSITIVITY

$$= \frac{A}{A + C} \%$$

Fig. 6.iii.



SPECIFICITY

$$+ \frac{D}{D + B} \%$$

Fig. 6.iv.

had reported to the University Health Centre for a follow up examination. There were two objections to this: first, not all the students referred by the screening receive their follow up examinations at the Health Centre - some consult private opticians, and some do not wish to receive attention at all. Second, and perhaps more important, these would find only the true referral and the over-referral groups, the other two groups being missed. Blumberg (1961) discussed studies of this type:-

> "In some studies only persons who are positive to the screening test are given a diagnostic evaluation. Some of these positive screenees are found to be free of the disease at diagnosis, but without diagnosis it is not possible to classify persons who are negative to the test; therefore cells C and D of the fourfold classification are indeterminate. If the assumption is made that all persons who are negative to the test are free of the disease, all of these cases will fall into cell D, and cell C will be empty. In this instance, sensitivity will be falsely stated to be 100 per cent, and the specificity of the test will be overstated. A study of this type is obviously of little or no value."

In view of this conclusion this approach to the problem was abandoned.

Evaluation of efficiency by questionnaire

The second follow up method considered was to send questionnaires to a random sample of students who had

previously been screened by the University screening method. It was hoped that if questions were put regarding whether they had had an eye test since the screening which resulted in the prescribing of ocular corrections or treatment for an ocular defect, or whether they had had any symptoms since the screening, then a reasonable idea of the success of the screening would be obtained. This procedure was adopted, there being no objections found at the time.

Data handling

The original screening results had been transferred to punched cards to facilitate their analysis. The questionnaire was, therefore, designed with a view to its responses being put into code form on a punched card to promote their easy comparison with the screening results. A copy of the questionnaire forms appendix III.

The population

The population from which the sample was to be drawn consisted of all those people who had been screened by the University screening technique since it was first used.

The sample

The sample had to be taken from one of the three undergraduate years still at the University, since it was impracticable to contemplate tracing the necessary number of postgraduates.

Of the undergraduate years, insufficient time had elapsed since the first year's medical examination for the treatment of all their referrals to have been concluded. The third year students were involved with their final examinations and it was thought that this would make the number of returned questionnaires very low, also it seemed unfair to add to their burdens. Therefore, the second year students were chosen. Another factor affecting this decision was that, although two years had elapsed since their

test, it did not seem unreasonable to expect the students to recall with some accuracy any treatment or symptoms they may have had during that time. It was also a sufficient time lapse for the development of symptoms which may have been missed at the screening.

Size of sample

It was decided that a sample size of two hundred would be large enough to make an accurate assessment of the phi coefficient (Φ). Assuming a 50 per cent response, an assessment based on previous questionnaire experience with students at the University, 400 questionnaires had to be sent out. In fact 404 questionnaires were distributed to 25 per cent of the second year students. The sample was selected by using random number tables from a list of second year students provided by the University Registry.

Response

227 questionnaires were returned, showing a response of 56.2 per cent slightly better than had been anticipated. Five of these replies were from students who had not received a screening examination on entry to University.

Problems of Definition

The classic definitions of the terms A : B : C and D used in the two by two matrix, require that the opinion of a practitioner be obtained for each student, his being the authority which decides whether the screening decisions were correct or not. His results are used as the datum from which to draw inferences about the efficiency of the screening method. His is assumed to be completely efficient ($\Phi = 1$). In this case, however, the only follow up information available

to us was that supplied by the student himself, via the questionnaire, and it was from this alone that conclusions could be drawn. The questionnaire was designed in the early stages of the investigation, and its limitations in terms of valid conclusions were not fully appreciated. These became apparent when the returns were being analysed. The definitions of the four groups had to be modified in view of the type of information which the questionnaire provided. Hence a true referral (A) was a student, referred by the screening, who reported that he had undergone a course of treatment or change of prescription. An over-referral (B) was a student who reported no action of this sort. A student who passed the screening, was considered to be an under-referral (C) if he reported that he was prompted by his ocular symptoms to seek professional advice. A correct non-referral (D) was a student who passed the screening and reported no course of treatment or change of prescription since the screening. These were the definitions which came nearest to the classic definitions given on p.24 . However, on analysis it was realised that they were inadequate. They became a major source of error in the investigation. These and other error factors are discussed in the following section.

Sources of error

There were two sources of error; those inherent in all questionnaire analyses, and those particular to the present study. The "inherent" errors were clearly illustrated by this investigation. The validity and extent of the information gained from the students was limited. Firstly it was dependent upon the student's ability to answer the

question, i.e. he had to have the knowledge required to provide an answer. The lack of understanding of any treatment he may have received would mean that he based his answers on his own conception of what occurred. Secondly, he may not have had accurate recall of events as they occurred, or when they occurred. Thirdly, the validity of his answers was dependant upon his clear understanding of the questions on the questionnaire, and his interpretation of them in the way intended. If he misunderstood a question he would give the answer to the question as he understood it and it could only, therefore, be misleading. He may not have considered a particular fact to be relevant, either out of ignorance, or because of some "mental reservations" regarding it. For example, the desire to appear either "normal" or "abnormal", and "special". Knowing the aims of the questionnaire may also have influenced his answers. Finally, error also arises in the interpretation of the student's response, which is particularly difficult when the interrogation is conducted by post, as tones and inflections of speech are not transmitted. Postal questionnaires have a further source of error in their poor response rate, in order to overcome the bias which this introduces to the sample, certain assumptions must be made and these assumptions are another seat of error.

Apart from these inherent errors, further error was introduced into this questionnaire because it asked questions using vague terms like "headache" and "eye strain" which were prone to subjective interpretation and which were completely unqualified by any degree of measurement.

Due to the "strained" definitions which had to be adopted there were certain situations in which a student could be wrongly classified. The following are some examples of these situations:-

(a) A student with one amblyopic (lazy) eye will be referred and examined by a practitioner to ensure that it is not due to a progressive condition. He may not, however, receive treatment or a change of prescription. Therefore, a perfectly valid referral would be classified as an over-referral because the student reports no treatment.

(b) A referred student who does not follow up the referral by seeking professional advice would be classified as an over-referral as he also would report no treatment, although his referral may have been quite justifiable. The reasons that a person may have for not obtaining professional advice are many and various, none of which would make the referral less valid.

(c) A student who passed the screening was considered to be under-referred if at some later date he is found to require treatment or a change of prescription. At the time of the screening he may have been correctly passed (non-referred). This objection can be made of any study which makes a comparison between the screening and follow up examination, when there is a large time gap between them.

Error detected

It was found that some of the students who claimed not to have had an eye test since the screening, also claimed to have had a change of prescription. These answers were incompatible. The error was probably due to a misunderstanding of the questionnaire. Although the number of students making this claim was small, it raised the question of the reliability of other claims made by the students, such as those who had been examined and reported a change in refraction, when in fact there was no such change. These claims could not be checked from the information provided.

Results

After analysing the returned questionnaires the following results were obtained:-

i. Referrals

In all, 37 of the 223 students screened were referred (16.3%), 18 of these to the University Health Service and 19 to an outside optician of their choice.

Of the 18 referred to the University Health Service, four did not keep their appointments, two went to outside opticians instead, and one went to both a private practitioner and to the University Health Service. The rest attended the Health Service as arranged.

Of the 19 referred to their own optician, four did not attend, one went to the University Health Centre, and the rest attended as advised.

Hence eight out of the referred 37 (21.6%) did not give themselves the chance of receiving further treatment.

ii. Group A - True referrals

Of the 37 students referred, 20 were given a change of prescription, and two received treatment (one for corneal irregularity, and one was given orthoptics).

Total = 22.

iii. Group B - Over-referrals

There were eight students who did not have an eye examination despite referral, and seven who had an eye examination but did not receive treatment or a prescription change.

Total = 15.

iv. Group C - Under-referrals

There were 42 students who passed the screening but later had a change of prescription, and five who received treatment.

Total = 47

v. Group D - True passes

Of the 186 students who passed the screening, 129 had no further examinations, and of the 57 who did have an eye examination, 10 did not receive treatment or a change of prescription.

Total - 139.

Assessment of efficiency

These figures were inserted into the 2 x 2 matrix (table 6.i.) and the phi coefficient of the screening found.

	Rx. or treatment	No Rx. or treatment	
Referred by screening	22	15	37
Not referred by screening	47	139	186
	69	154	223

Table 6.ii.

The phi coefficient of correlation Φ

$$\frac{(22 \times 139) - (15 \times 47)}{\sqrt{37 \times 186 \times 69 \times 154}}$$

$$= +0.3$$

Conclusions on the estimation of efficiency found by questionnaire

The phi coefficient obtained from this experiment, to represent the efficiency of the University visual screening technique was +0.3. This is very low, less than that achieved by the "teacher observation" technique used as a screening method in the Orinda study. Blum et al of the Orinda Study considered that teacher observation did not "merit consideration as a screening method". Therefore, either the screening technique employed

by the University is in fact very inefficient, or the method employed to measure its efficiency is very unreliable. The evidence already presented hereand later suggests that the latter is almost certainly the case. This suggested that a more controlled experiment should be carried out to make a more accurate assessment of the screening technique.

Perhaps the most useful contribution made by this experiment was in drawing attention to the apparently large proportion of students (over 20% of those referred) who, although referred by the screening did not follow this up by a clinical examination. The figures quoted may be very inaccurate, but it is disturbing to find any at all. It, therefore, seems that an important aspect of making a screening system efficient is to carefully follow up the referred cases.

2. PRECISION

Whilst the investigation described previously was being carried out, the second aspect of the screening assessment was being considered; namely an estimate of the precision or repeatability of the University screening technique.

Precision refers to the ability of a test to give consistent results in repeated trials. In order to assess the precision of a test the subjects must be re-tested by the same screening test and the two (or more) results compared. The extent of agreement between the two results is affected by many variables. These variables are levels of illumination employed, skill of the tester, position of apparatus, subject motivation etc. It is impossible to control all these factors, particularly those involving people. However, if as many as possible of these variables are kept constant, then it should be

possible to measure the maximum precision obtainable with the screening method. Assuming that, at the other end of the scale, it is possible to have little or no agreement between the two results, then the upper and lower limits of the extent of agreement can be specified. The agreement to be expected in normal circumstances must fall within the two limits.

Precision measured by re-screening a group of students

A seemingly useful test-retest situation was that presented by the "Education Students". This group of students were primarily screened as undergraduates on entering the University of Birmingham. Three years later they enrolled on the Education Course at the University and because of this had to be re-screened. After collecting the results of the two screenings, objections as to the reliability of using this situation were raised. These objections were as follows:-

1. The sample of students was not random as it represented an academically successful group.
2. The screening was carried out with an interval of three years between trials.
3. The location of the screening was different, and so extraneous factors such as illumination levels were not constant.
4. A different examiner was used in the second screening.
5. The second examiner knew that the re-test evaluation was being made, and had knowledge of the results of the first screening.
6. The results of the first screening effected the results of the second, because a subject failing the first screening would probably have received treatment in the interim.

The validity of these objections led to the abandonment of this attempt to specify the precision of the screening technique.

The alternative to this was to carry out a more controlled test-retest experiment.

7. A CONTROLLED EXPERIMENT TO FIND THE EFFICIENCY AND PRECISION OF FOUR VISION SCREENING SYSTEMS

The evaluation of screening efficiency made by sending follow up questionnaires to the students who had been screened was found to be very inaccurate. A more controlled experiment to assess screening efficiency was found to be necessary. Similarly, when trying to assess the precision or repeatability of the University screening technique, it was found that too many factors were permitted to vary at once,and there was little control over the test-retest situations.

To overcome these difficulties, an experiment was planned whereby a sample of subjects were to be screened under controlled conditions, and then given a full clinical examination. A proportion of this sample was to be re-screened so that the screening precision could be found.

As this involved the subjects in having a full refraction it seemed economical to apply the information from it as widely as possible, by comparing its results not only with the University Visual Screening results, but with those from other screening techniques as well. Four separate screening systems were available at the Department of Ophthalmic Optics, University of Aston in Birmingham. These were consequently included in the experiment. The main advantage of doing this was that the efficiency and precision of the University Screening Technique could be compared with those found for other widely accepted screening techniques, under truly comparable conditions.

Aims of the experiment

Primarily the experiment had two main objectives;-

1. To find the phi coefficient of each screening technique under consideration.

2. To find the precision or repeatability of the screening techniques.

There was, however, a secondary investigation planned, into the factors which affect the efficiency and precision of the screening tests. This could be divided into three parts:-

a. To find the accuracy of the major subtests within the screening tests.

b. To find the sensitivity and specificity of the screening tests and to discover how these varied when the referral cut-off points were changed.

c. To find the "cost" of the various techniques in terms of time, personnel, and money.

The screening methods studied

The four screening systems studied were as follows:-

The MODIFIED ORTHORATER, manufactured in the United States by Bausch and Lomb, is a portable, internally illuminated apparatus. The twelve tests are presented as stereoscope slides. The slides are transilluminated and placed in position manually.

The MASTER VISION SCREENER or "MAVIS" is manufactured in Great Britain by J. & R. Fleming Ltd. There are fourteen test cards, produced photographically, and mounted on a rotating drum. The cards are illuminated from the side and the drum is rotated by hand.

The TITMUS OPTICAL COMPANY PROFESSIONAL VISION TESTER is also a stereoscopic instrument, produced in the United States. Twelve tests are presented to the subject automatically (by a push-button control held by the operator). The slides are transilluminated.

The UNIVERSITY VISION SCREENING METHOD was the fourth technique analysed. It has been described previously (p.19). However, instead of using one examiner per sub-test, one examiner performed all the sub-tests.

Choice of sample

The choice of sample had to be related to the population to which the screening was normally administered, namely the first year intake of students to the University of Birmingham The most readily available sample (students of the University of Aston in Birmingham) was probably fairly representative of this age group and ability level.

This was a non-homogenous group, the subjects having different motives for volunteering. The subjects were obtained by advertising for volunteers within the University. A nominal fee was offered and the major incentive in the majority of cases was, therefore, financial.

100 subjects were used.

Subject controls

Each subject was given standardised instructions, standard monetary payment and promised knowledge of results. All four screening tests were given to each subject. An "incomplete block" experimental design, where each subject is tested on only two or three of the screeners, would have made the experiment less tiring for the subjects, but the experiment would have become more lengthy and a much larger number of subjects would have been required. The order in which the screeners were presented to the subjects was determined by a Latin Square design, this ensured that subject factors such as learning, adaptation, fatigue etc. were randomised over all the screeners. Using the same subjects on all

the equipment ensured that individual subject differences were kept constant for all techniques.

The subjects were told that they were not competing against the clock, but that undue delay was to be avoided. They were also told that the object of the experiment was to "compare different types of screeners".

Screener controls

When making qualitative comparisons about complex systems, the systems must be working under optimum conditions. The manufacturers of each screening instrument recommend the cut-off standards to be used for subjects from various occupations. For example, inspection and close machine work, operation of mobile equipment, machine operation, unskilled labouring etc. The standards adopted for this experiment were those recommended for "clerical workers" as this seemed the occupation most nearly related to that of a student. They were taken to be the optimum standards for referral. The manufacturers operating instructions were followed carefully.

The screening examinations were carried out in a room which was illuminated solely by overhead, fluorescent strip lighting.

The same examiner was used to administer all the screening tests. This introduced constant errors into each one. However, overall effects such as operator fatigue, and the application of knowledge obtained about the subject in one technique, to interpretation of the results of another technique, was randomised by the Latin Square design.

The Clinical examination

The clinical examinations were carried out by an Ophthalmic Optician. A single consulting room was used, therefore systematic errors were introduced both from the apparatus and from the optician. This could only be avoided by using a large number of examiners in various situations, and this was not practical. It can be stated that the clinical situation was constant for all subjects examined.

The opinion of the practitioner and the results of her examinations were taken as the standard against which the screening techniques were compared. It is, therefore, more correct to use the phi coefficient as a measure of relative efficiency, rather than absolute efficiency of each technique.

The subjects were given the four screening tests first, following by the clinical examination, the whole routine taking one hour per subject. Neither the subjects nor the optician were told the results of the screenings until after the experiment.

There was one trial run before the beginning of the experiment proper, to finalise the routine and for the examiner to become accustomed to the apparatus.

Measurement

The measurement required to find the phi coefficient for each screener was "pass" or "fail" by the screening, and "requires professional treatment and/or advice" or "does not require professional treatment and/or advice" from the clinical examination.

Re-test extension of the experiment

A sample of the subjects screened in the first part of the experiment were re-screened at a later date. The number of students re-tested in this way was eight.

The screening section of the examination was repeated using as far as possible the same conditions as previously.

Results

Efficiency of the techniques

Results were placed on the 2 x 2 matrix (as in table 6.i) to enable phi to be calculated for each method.

	Orthorater		"MAVIS"		Titmus Opt.Co.Scr.		University technique	
	Not Ref.	Ref	Not Ref.	Ref	Not Ref.	Ref	Not Ref.	Ref
Does not require prof. treatment.	45	24	45	24	43	26	57	12
Requires prof. treatment	14	17	16	15	9	22	9	22
Phi coeff. Φ	0.19		0.13		0.31		0.52	

Table 7 i

Discussion of results

The phi coefficients found by this method are low, but not unexpectedly so. Previous studies of this kind have also found little relationship between screening results and clinical findings. The following list gives the results of major screening studies:-

St. Louis Study,	Orthorater	$\Phi = 0.33$
(1952)	Snellen Types	$\Phi = 0.42$
Robinson, (1953).	Orthorater	$\Phi = 0.7$
	M.V.T.	$\Phi = 0.65$
Orinda Study	M.C.T.	$\Phi = 0.95$
(1956 results)	M.V.T.	$\Phi = 0.59$
	Telebinocular	$\Phi = 0.57$
	Cal.State reg.proc.	$\Phi = 0.41$
	Nurse observation	$\Phi = 0.4$
	Teacher observation	$\Phi = 0.24$

The notable exception to these low phi values is that for the Modified Clinical Technique of the Orinda Study.

The measurements are very varied. For example, the phi coefficient has been found for the Orthorater, to vary between 0.19 (in the present study) and 0.7 (found by Robinson in 1953). As the tests were applied to different populations, by different testers and judged against the standards of different professional examiners, the comparisons are not really valid.

The conclusion at this stage of the investigation is the same as that of the St. Louis Study: "It is evident from the data reported that none of the vision testing methods studied provide more than a rough screening procedure". The more important question is raised of why this should be so.

Precision of the techniques

Precision refers to the ability of a test to give consistent results in repeated trials. The pairs of results obtained by re-screening eight of the subjects were compared in order to measure the precision of the four screening methods.

In view of the small sample, and because the variables were measured mainly on ordinal scales, it was not possible to make the comparison using a powerful, parametric statistical test. Instead the non-parametric Kolmogorov-Smirnov two-sample test was employed. This two-tailed test is sensitive to any kind of difference in the distributions from which the two sets of measurements were drawn.

It was found that there was no significant difference between the distributions of the two test results, for any of the screening sub-tests. The screening methods are, therefore, acceptably precise.

8. Factors affecting the efficiency of visual screening techniques

The errors involved in visual screening were discussed in section 6.1. To recapitulate briefly they are due to the fact that:-

i. The number of visual characteristics tested is restricted by the time available, and the experience of the tester and subject.

ii. The sub-tests themselves may be inaccurate.

iii. The pass/fail cut-off points may not be at the optimal positions.

iv. The interpretation of the subject's responses may be erroneous.

This section discusses these error factors with reference to the screening techniques examined previously, and the population to be screened.

8.i. Visual characteristics tested

To decide which visual characteristics should be tested by the screening, the nature of the population to be screened, with respect to its age and occupation, must be considered, together with the reason why the screening is to be performed.

For example, with regard to occupation: textile workers, electricians, train drivers, and all persons whose occupation involves the correct recognition of colours, should have their colour discrimination tested. This is not necessary for all occupations. Similarly, crane drivers should be tested for stereopsis, and inspection workers should have their acuity at the inspection distance checked. Therefore, when screening

populations from different occupations the emphasis of visual characteristics to be tested changes.

Similarly the age of the population to be screened influences the choice of visual characteristics to be tested. For example, glaucoma is a condition which mostly affects people over the age of thirty, and so it need not be included in the list of characteristics which are tested in a younger population. In young children, any disorder which prevents the normal development of binocular vision must be found and treated at once, but in an older population, whose binocular state is well established, these disorders are less important.

Here it is useful to discover what other workers in the field of visual screening have considered to be the most important visual characteristics of the populations in which they were interested.

Diskan in 1955, and again in 1963, stated that visual acuity, excessive manifest hypermetropia and eye muscle balance were chosen as those visual functions most important to a school child. Similar statements are made by Reese in 1964, and Blackhurst and Radke, also in 1964, although there is a slight disagreement as to whether the muscle balance at near should also be included. Gallager and Gallager (1964) suggest that early colour vision testing should be carried out in children using H.R.R. plates. The Orinda study, suggests a far larger range of tests for school children: ocular dominance, near cover test, near point of convergence, visual acuity, distance skiametry, and inspection of optical media and external tissues. Many of these tests cannot be performed

by an untrained person, and so if the resources of the screening administrators are not great enough to employ a qualified practitioner, these tests cannot be included in the test battery.

An article in Nature, 1965, states that monocular distance and near acuity tests are not sufficient for testing drivers vision, it suggests that they also be given binocular vision tests and visual field tests.

"If all applicants over forty five years old, and pay-roll employees over forty five, in all industries were given this simple test periodically (yearly), two to three per cent of this population having undetected glaucoma would be uncovered." This was a statement made by Kuhn in 1957, and the simple test which he mentions is the Harrington Flocks Visual Field Screener.

Powel et all (1964) suggest that the following tests be included in all screening programmes: visual acuity, muscle balance (near), hypermetropia (for seven to eighteen year olds only), stereopsis, colour perception, near visual acuity (tested in people over thirty five), similarly visual field tests for that age group only.

Despite the diversity of tests that can be applied "there is good agreement that the most important function to be tested is central vision". This is the opinion of Sloane and Savitz (1963), and it is an opinion expressed earlier by Hatharway in 1959, who calls it "the most important single test", which according to the National Medical Foundation for Eye Care "identifies more children requiring eye care than any other single test". Diskan (1963) categorically states that "There is no substitute for the visual acuity test. It always should be done."

There is one overriding justification for performing a screening test, according to Schwarz (1965) and that is that "the condition discovered be amenable to treatment". If this is accepted then there is no justification for testing colour discrimination, unless the term"treatment" can include "advice on vocational choice", and the possible prescription of colour filters which may assist in colour discrimination.

Nature of the population to be screened at the University of Birmingham

The general principles involved in selecting the set of most important visual characteristics will now be applied to the University of Birmingham screening situation.

The University population consists of freshmen at the University of Birmingham, their average age being 19.7 years (a figure found from the 1967 sample). They belong to all faculties, and their ultimate occupations will be varied. Their immediate occupation, and the one which they have in common, is that of student.

The purpose of the screening as described by the University Medical Officer (Bolton 1955) is the identification of students who need further examination, this being based on their liability to eye trouble during their University course. The students will be liable to eye trouble if they do not possess the visual skills required by their occupation, or if they have some pathological condition affecting their eyes. The visual skills required by students are: good distance acuity to enable them to see clearly the visual aids used in lectures, and good near vision and the ability to use it for long periods without strain.

Refractive errors

The majority of refractive defects found will be small as any student who already possesses an optical correction will wear it during the screening tests. Hence those errors of refraction which were readily detectable at an early age have probably already been corrected. In the 1967 screening, 46% of the students wore some optical correction, a similar proportion was found in the experiment performed to find the efficiency of the screening. The distribution of uncorrected errors found in that experiment is shown in figs. 8.i. and 8.ii.

These smaller defects cause only small inconvenience, which was easily overcome in the school situation. For example, myopes could sit near the front of the classroom in order to read from the blackboard. Young hypermetropes, asymptomatic except when undertaking prolonged close work, probably experienced difficulty at examination time only. Similarly, astigmats who overcome their error by continual changes of accommodation, or adoption of a compensatory head posture, probably only experience difficulty during long periods of critical viewing tasks.

As a student is required to undertake long periods of reading and close work, then any refractive error which can be overcome by accommodative effort may produce stress. It is, therefore, desirable to correct small errors of refraction.

Binocular anomalies

Any student exhibiting a complete absence of binocular vision should be referred to ascertain whether it is due to a progressive pathological condition, and to instigate any treatment necessary.

Heterophoria is a very prevalent condition, but does not normally give rise to symptoms of discomfort if the student has good fusional reserves. Gibson, in his book on Orthoptics (1955)

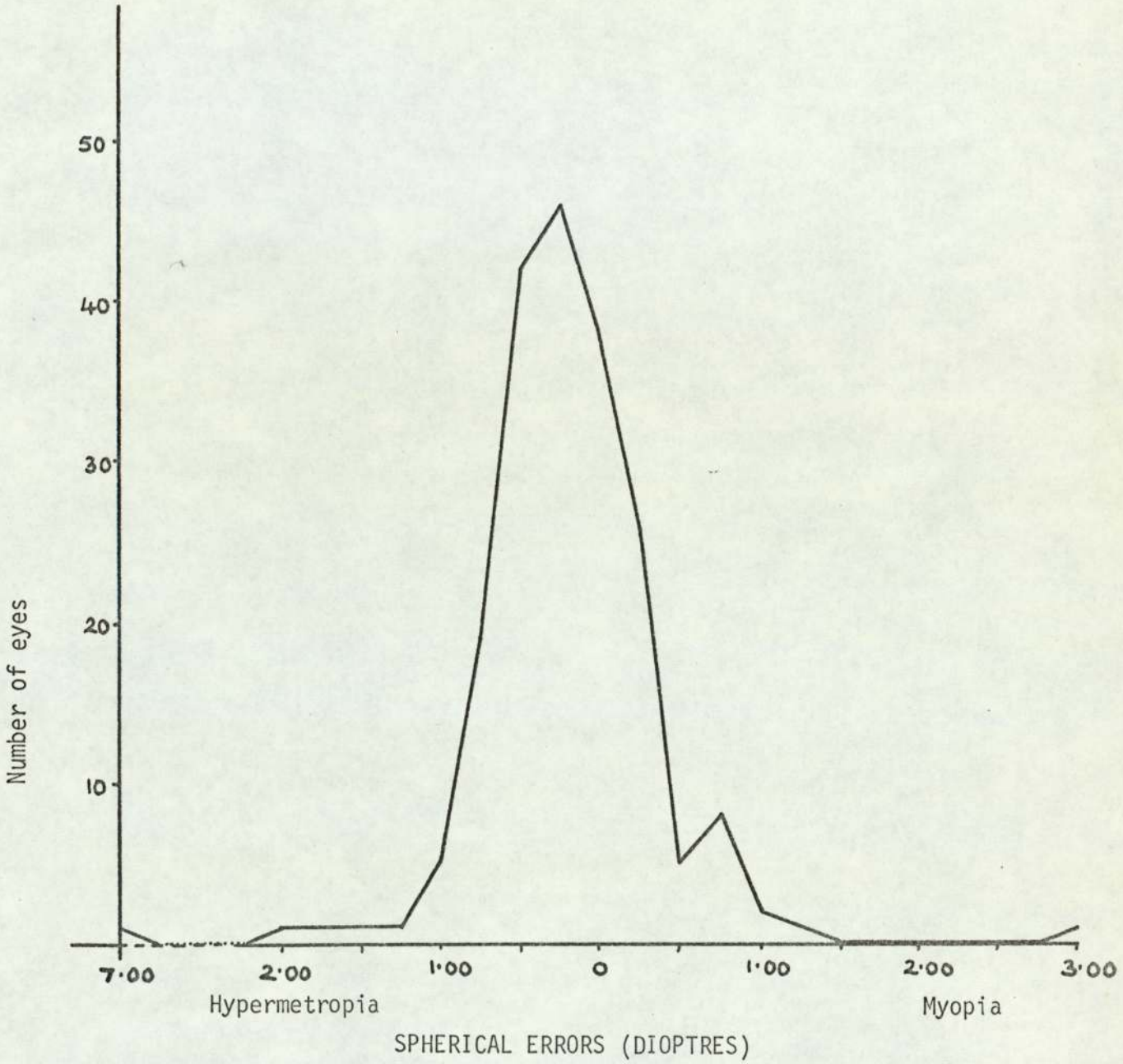


FIG. 8 i. Distribution curve of the spherical component of refractive errors found in the experimental sample.

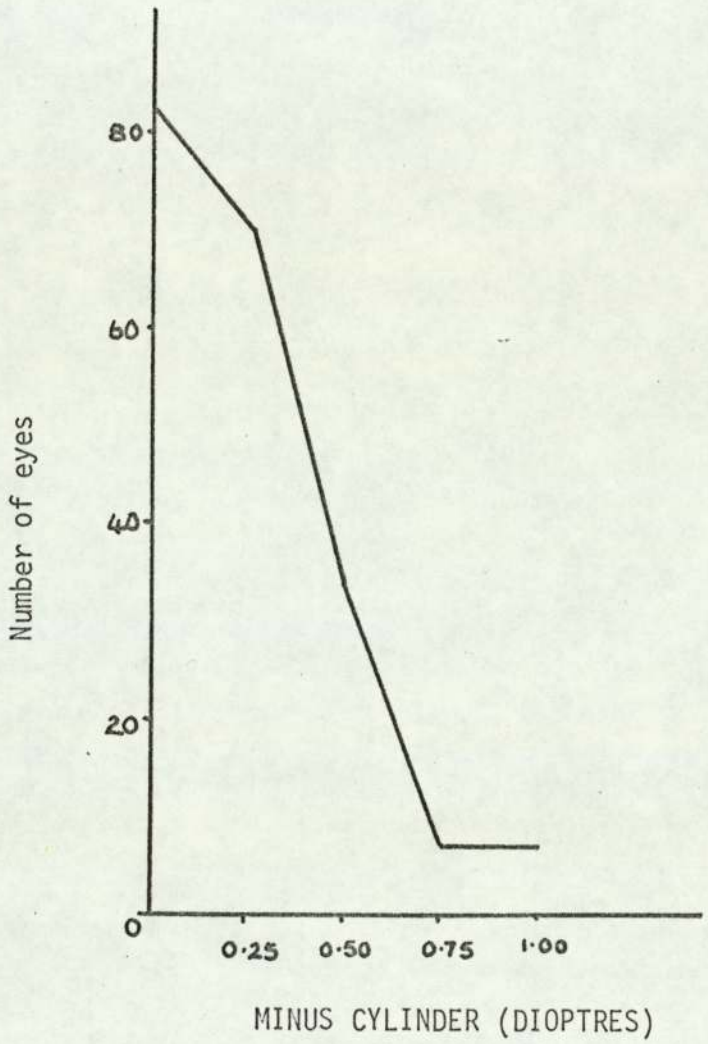


Fig. 8 ii. Distribution curve of the cylindrical component of refractive error found in the experimental sample.

mentions some conditions in which heterophoria gives rise to symptoms:

"a. an inability of the fusional reserves to cope with the amount of heterophoria.

b. abnormal demands on accommodation, convergence or both.

c. poor physical health, and

d. in the psychological make up,for in the happy individual, devoid of worry with few demands on physical and mental reserves, symptoms rarely occur."

A student's life produces most, if not all of these conditions, and it is, therefore, very probable that a student with poor fusional reserves will experience symptoms.

Pathological ophthalmic conditions

Of all the pathological conditions, the following are most likely to be found in the University population; congenital pareses of an extra-ocular muscle producing an incommitant strabismus. As strabismus was found in only 1.6% of the 1967 sample, and the majority of these were concomittant, then the incidence of paretic muscles is low in the student population. Defects of colour vision are prevalent among the male University population, as in the general population. They were found in 9% of the males of the 1967 sample. Most of these students were already aware of their defect. It should only be necessary to test the colour vision in the male students whose vocation requires them to have good colour discrimination. Inflammatory conditions of the eyelids and conjunctiva are common; however they are painful and

irritating to the student and he is usually aware of their presence. Other pathological conditions which could effect this population have a low prevalence, which means that if special tests are included in the screening to discover these conditions, then the number of false positives will be greatly enlarged.

It has been shown (Thorner and Remein, 1961) that when the prevalence of a disease is low (in the order of 1 - 2 per cent) most of the population will be free of the disease and the positive results, even for a highly sensitive and specific test, will include a large proportion of false positives. If the prevalence is increased then there is a negligible reduction in the number of false positives but the number of true positives increases in proportion to the increase in prevalence. Therefore, the proportion of false positives is reduced.

Therefore, if screening can be directed towards a high prevalence group it is more successful. Hence the screening becomes a two-stage process in which selection of a high prevalence group is the first stage.

Pathological ophthalmic conditions tend to produce reduced visual acuity, binocular imbalance, and occasionally changes in colour perception. As all these conditions are to be tested in the screening, the tests could be considered as the first stage in screening for disease, in selecting a high prevalence group. Hence those people who are failed by those tests may be re-screened for specific pathological conditions, as the second stage of screening for disease.

Choice of sub-tests used to measure the visual characteristics

Having decided which characteristics should be tested by a screening battery, the most suitable methods of testing them must be found.

Lippman in 1962, suggested that ideal screening procedures should be "simple, fast, inexpensive, valid, reliable and productive". He neglects to mention that a screening procedure which is ideal in one situation is not necessarily ideal in another context. For example, the Orthorater, which has been used fairly successfully in many industrial situations in the United States and Great Britain was evaluated in a pilot survey by Argarwal and Das (1964), who were considering it for use in screening Indian railway workers. Because of illiteracy among the population being screened many subjects took as long as thirty minutes to complete the test battery, and the results, according to the report, often strayed "far from accuracy". Similarly a test suitable for an adult population may be impossible to apply to a population of young children. Constraints are also placed upon the choice of tests to be included by the situation in which the screening must be performed, the main limiting factors being the space and time available for the tests, the level of illumination available and its variability, and finally the experience and abilities of the examiners.

Time available

Students must pass through the "stations" at the rate of one per minute. Hence, from the start to finish of the ophthalmic tests each student takes six minutes.

There are two ways of achieving this: either the method of presentation remains similar to the one used at present i.e.

one examiner per sub-test; or one examiner takes one subject through the whole series of tests, there being several examiners working simultaneously, i.e. the method used in the G.O.C. caravan containing "Mavises" each of which provides the entire screening for one subject.

Of these two methods the first is better in that only one set of tests is required, and this cuts down the initial cost.

Space and illumination

The space available is an area about 10' x 10'. Illumination is poor, relying mainly on daylight, and so it is not constant. Some local illumination is available.

Examiners

In past years examinations have been performed by final year ophthalmic optics students, with an ophthalmic optician making the assessments.

The population

The population to be screened is literate in the English language, and intelligent. Therefore, fairly complex tests may be used, provided that it is remembered that the subjects are not trained observers. With these criteria in mind an analysis was made of tests which could be included in the screening.

A. Tests for ametropia

Here either an objective test such as retinoscopy may be used or a subjective one involving a chart used in conjunction with trial lenses.

Retinoscopy has been used as a screening examination, by Hirsch (1950) and the Orinda Study. Hirsch examined children and found the main problems to be keeping the child's attention,

yet maintaining a relaxed accommodation without cycloplegia. A further study by Hirsch demonstrates that two determinations of refractive error made on the same individual may differ by as much as 0.75 dioptres. The Orinda Study, however, recommends retinoscopy as a screening method provided that the subject looks through +1.50D lenses for at least a minute before the commencement of retinoscopy, so that his accommodation is relaxed.

There are several objections to using retinoscopy in the University screening situation, apart from its doubtful reliability. Firstly, retinoscopy requires constant low levels of illumination, without glare from extraneous light sources. This is difficult to achieve in an area where many other people are working. Secondly, retinoscopy takes several minutes, even when only an estimate of ametropia is required, particularly as some time has to be spent relaxing the subject's accommodation. Finally, and most important, a skilled examiner is required. A basic precept of screening is that a non-skilled person can administer the tests. For these reasons retinoscopy is rejected as a screening test.

Ametropia must, therefore, be measured by a subjective method. This necessitates designing a chart and deciding on the best method of presentation. The main criteria of a test type is that it is standardised, and its limitations known. A comprehensive review of test types was made by the British Standards Institute Sub-committee on Ophthalmic Test Types (Bennett 1965). From this review they decided on a British Standard Chart. The B.S. Chart (see B.S. 4274 : 1967) has adopted English style non-serif capital letters, and includes only those letters of

equal legibility, namely D, E, F, H, N, P, R, U, V, and Z, the ratio of letter height to width being 5 : 4. A chart following the British Standard recommendation should, therefore, be used to find the subject's V.A. in a screening situation. There are several advantages of using letters as a screening test; firstly they require little explanation, and so errors in understanding are avoided; secondly when the examiner has remembered the correct pattern of letters, mistakes are easily noticed, even when the examiner's attention wanders, as it probably will after a few hours of repetitive work; finally, the V.A. test can be applied by an unskilled examiner.

As the illumination available is variable and not conforming with the recommended levels, the chart should be presented within an internally illuminated box. It would be impractical to have a box as long as the standard testing distance i.e. 6 metres; a small box should, therefore, be used with the chart presented at "infinity" by employing an optical system. This may introduce accommodative errors but they probably have less effect than the errors which would be introduced by poor and inconsistent chart illumination.

As has already been mentioned, distance visual acuity tests do not discover low hypermetropia. Hence a second test should be added to the measurement of distance acuity, that is the acuity achieved when viewing through a pair of spherical plus lenses. The power of these lenses is determined by the amount of hypermetropia to be detected.

BINOCULAR ANOMALIES

Strabismus is usually associated with depressed acuity in one eye. This will be found by the distance visual acuity test. The presence of suppression will also be evident in any binocular test.

The detection of poor fusional reserves is more difficult. The Mallet Test discovers the presence of retinal slip, which occurs when fusional reserves are low, and a phoria is not compensated. However, this is a difficult test to perform because it requires careful observation by the subject. In theory the Mallet Test would be very useful, but because of the limitations imposed by the skilled observation and interpretation required it cannot be used in screening.

The most common method of finding non-compensated phorias is to assume that large phorias are less likely to be compensated than small ones, and to select those people with large heterophorias. In doing this the variation in tolerance to phorias between individuals must be ignored. In fact it is the tolerance of a phoria which should be measured, but here it must be assumed that the tolerance is directly proportional to the amount of heterophoria.

As the amount of heterophoria measured depends on the test object's distance from the eyes, then the test distance must be specified. This should be the distance at which the student will be doing the majority of his work, that is at reading distance.

Visual characteristics tested by existing University screening method

Screening batteries should have only a small number of sub-tests, because, as Lippmann stated in 1962, the more comprehensive a test battery the greater the over-referral rate. A detailed study of the 1967 screening results was made in order to discover whether any parts of the University screening test were redundant. The purpose of the investigation was to decide which combination of sub-tests measuring various visual characteristics, gave the best estimate of the refer/non-refer criterion. Hence the relative contributions of the tests in predicting the criterion, had to be assessed. The greater part of the prediction could probably be attributed to a relatively small number of tests, and the inclusion of additional tests probably contributed only a small amount to prediction.

Ideally, the tests for variables which show a high correlation with the pass/fail criterion, and a low correlation with other tests should be identified, when the assumption must be made that these tests measure different aspects of the criterion, and contribute greatly to prediction. Conversely, any test found to have a high correlation with another test, means that the inclusion of both tests, instead of one or the other, is unnecessary and contributes little to the prediction achieved.

Statistical test used

The aim of the investigation was to find the correlation between different sub-tests and between the individual sub-tests and the pass/fail criterion.

The University screening method records variables for each student on the following scales of measurement:-

	Variable	Scale of measurement
a.	Sex	Nominal
b.	Correction worn	Nominal
c.	Date last test	Ordinal
d.	Hypermetropia, L. eye.	Ordinal
e.	Hypermetropia, R. eye.	Ordinal
f.	V.A. L. eye.	Ordinal
g.	V.A. R. eye.	Ordinal
h.	Near acuity. R. eye.	Ordinal
i.	Near acuity. L. eye.	Ordinal
j.	Colour vision	Nominal
k.	Heterophoria measurement	Interval
l.	Assessment (criterion measure)	Nominal

Table 8 i.

It can be seen from Table 8 i that most of the variables are measured on ordinal scales, and some only on nominal scales. Only one correlation test can be applied to data of this sort, and that is the contingency test. However the contingency coefficients obtained from matrices of different sizes (as they would be in this case) cannot be directly compared. Hence another method of comparison had to be found. Instead of calculating the correlation between variables, the chi squared (χ^2) test was applied to the data, to decide whether samples from one variable could be said to have come from the same population with respect to the other variable.

The Null hypothesis

The null hypothesis was that the proportions of the classes of one variable were the same for all classes of

(a) the assessment criterion

(b) the other variables.

Data collection and manipulation

The data obtained from the 1967 visual screening of freshmen was collected, coded (using the code described in Appendix II) and transferred to punched cards. Records were obtained for the 1,525 students screened.

The cards were sorted to produce bi-variate distribution matrices from which chi square values could be calculated.

Level of significance

A probability of 0.01 was taken to be the level below which the null hypothesis is rejected.

Results

The probabilities associated with the χ^2 values were found.

(a) The variable and the assessment criterion

It was found that for the following variables, the proportions of the classes of those variables were the same for all classes of the assessment variable, that is they did not contribute to the subject passing or failing the complete screening:

Sex of student

The null hypothesis was rejected in the case of the following variables, which, therefore, must contribute towards predicting the assessment criteria:

Correction worn
Date last test
Uncorrected hypermetropia (L & R)
Visual acuity.

The tests which do not contribute towards prediction can be dropped from the screening.

(b) The variable and all other variables

Those tests which have their class proportions the same for all classes of the other variables are the ones whose results do not overlap with the results of other tests, and so are more predictive.

Variable	Number of other variables (maximum 10) with which the null hypothesis is rejected
Sex of student	2
Correction worn	5
Date last test	6
Uncorrected hypermetropia L	2
Uncorrected hypermetropia R	2
V.A. L	4
V.A. R	4
Near V.A. R	χ^2 not accurate
Near V.A. L	χ^2 not accurate
Heterophoria	χ^2 not accurate
Colour vision	1

Conclusions

It was found that the following tests measured different aspects of the pass/fail criterion, and contributed a major part to prediction of the pass/fail criterion. Null hypothesis is rejected with less than five of the other variables (five being an arbitrary number):

Sex of student
Uncorrected hypermetropia R & L
Visual acuity R & L
Colour vision.

8.ii. ACCURACY OF SUB-TESTS

The sub-tests of the four visual screening methods studied previously were analysed to find their accuracy when used to screen a sample of people similar to those of the University population.

Definition of Accuracy

"Accuracy is the ability of the test to give a true measurement of the item being tested" (Principles of Screening for Disease, 1961).

Finding the accuracy of sub-tests involves comparing the measurement made by the screener, in the screening situation, with the "true measurement" which is taken to be the measurement made by the professional person in the clinical situation.

To find the accuracy of the sub-tests by experiment

The experiment designed to find the efficiency of the four screening techniques has as one of its secondary objectives assessment of the accuracy of the screening sub-tests. The clinical measurement and screening measurement should ideally have been made in a randomised order on each subject, to randomise fatigue effects and other subject and examiner variables. However, as this would have made the experiment considerably longer and more complicated for a very small gain in its accuracy, the subjects were all screened first and given the clinical examination last. The screeners were applied in a randomised manner, according to a latin square design.

Measurements were made of several variables by each of the four screening methods and the variables were also measured under clinical conditions.

Stat. test used

A one way analysis of variance was used to discover whether there was any significant difference between means, of

the distributions found for each method of measurement. If a significant difference was found, the t-test was applied to compare each method of screening measurement with the clinical measurement, in order to discover which of the screening methods were significantly different from the clinical measurement.

RESULTS

DISTANCE VISUAL ACUITY TESTS

All four methods of screening measure visual acuity monocularly, by dissociating the eyes; however, the type of chart used is very different in each case - the Titmus Optical instrument uses Landolt rings, the Orthorater uses a "checker-board" target, and the "Mavis" uses words. In all of these cases the distance is simulated optically; however, in the University technique the distance is "real" and the chart consists of "Snellen" letters.

The scales on which the sizes of targets are based are different for each method of screening and they are not directly comparable. The Orthorater and Titmus Optical apparatus use a decimal graduation based on the decimal equivalent of the Snellen fraction: from 0.1 to 1.5 in 0.1 steps in the Orthorater, and from 0.1 to 1.4 in the Titmus Optical apparatus, also in scale divisions of 0.1. The "Mavis" uses a scale progression based directly on the Snellen Notation of the letter sizes, whereas the University technique uses only two letter sizes. The letter chart with which all of these systems were compared, the clinical chart, uses letters whose size graduation is determined by the Snellen Scale.

F-tests performed on the distributions of visual acuities for the right and left eyes (considered separately) showed a significant difference in distribution of acuities measured by

different screeners for the same population.

Assuming that the population variables have been successfully controlled, then the difference must be due either to the different modes of target presentation, different test types used, or the fact that the F-tests were comparing acuity scales which were not directly comparable (or a combination of these reasons).

The cumulative frequency curves of visual acuity (termed "more than" curves by Cutler and Davey 1965) plotted for each screening method and the clinical examination show that methods of screening which have similar acuity scales have similar distributions, e.g. the Orthorater and Titmus Optical Co. apparatus are similar, and the Mavis and University method are similar. Figs.8iii-v. To compare the screening methods in a more realistic way the acuity measurements were all modified to fit on to a single revised scale. This necessitated a loss of information from the scales which had small divisions, but produced a more genuinely comparable set of figures, which, when subjected to an F-test were not found to be significantly different for each apparatus in the right eye, but to be slightly significantly different in the left eye. When t-tests were performed comparing individual screening values with the clinical values for the visual acuity of left eyes, the difference was found to be due to an inaccuracy in the Titmus Optical instrument; the other screening methods being accurate, within the limits of the measurements made in this experiment.

D. A. Gordon et al in 1954 evaluated wall chart presentation and compared it with instrument, simulated distance. They report that the two methods of presentation were of equal

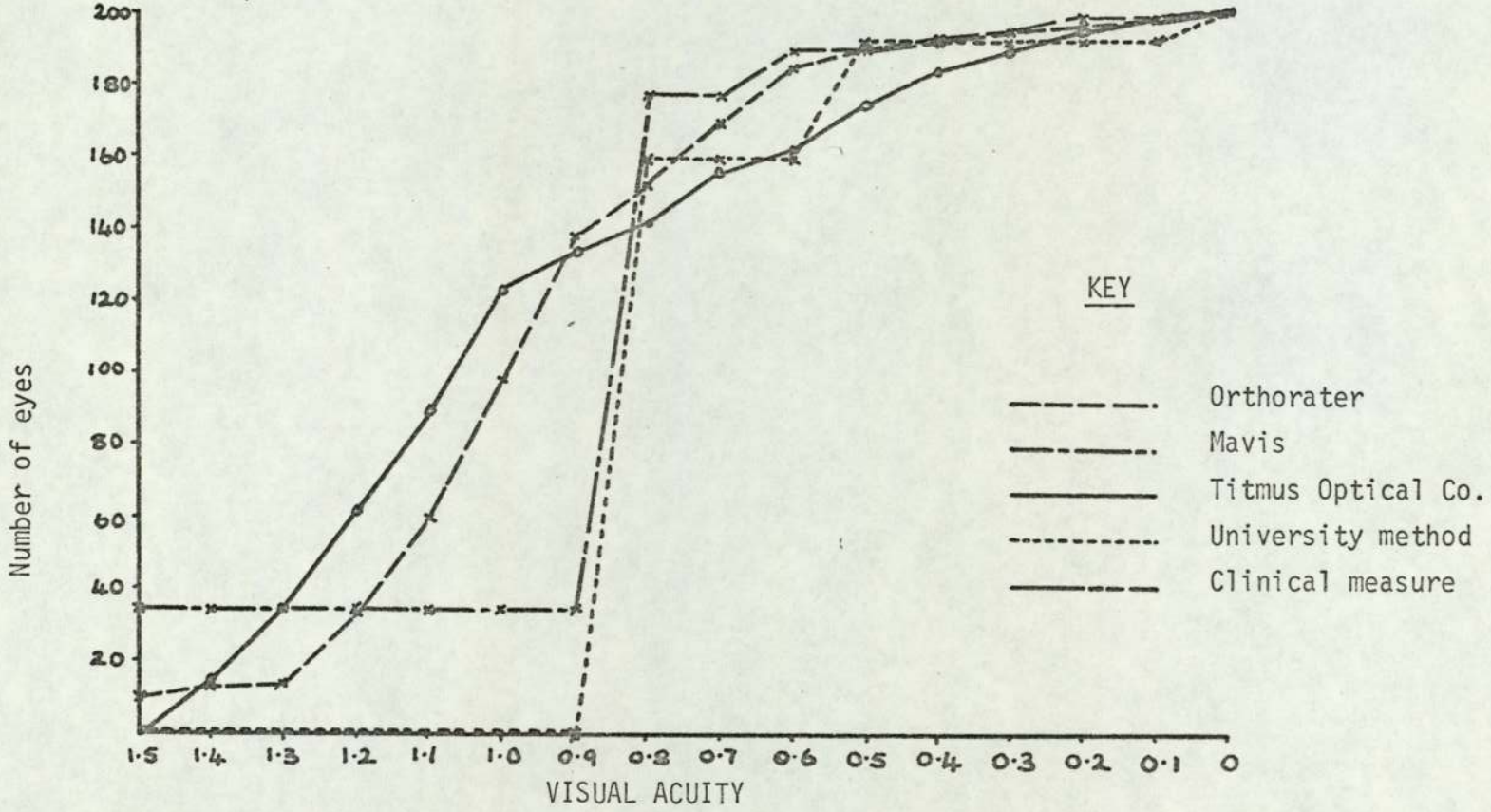


FIG. 8 iii. Cumulative frequency distributions of visual acuity (both eyes) measured clinically and by each screening method.

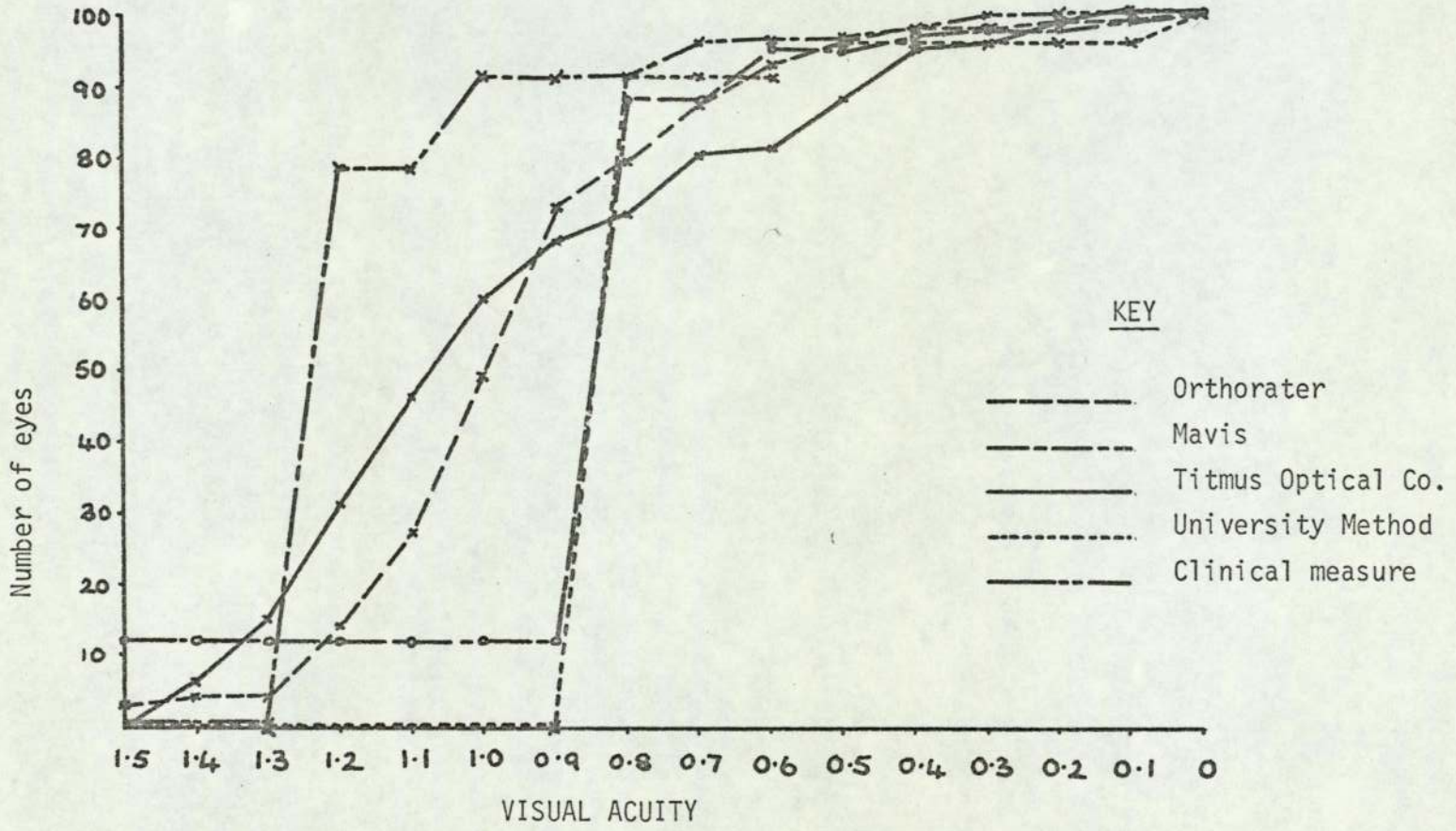


FIG. 8 iv. Cumulative frequency distributions of visual acuity (right eyes) measured clinically and by each screening method.

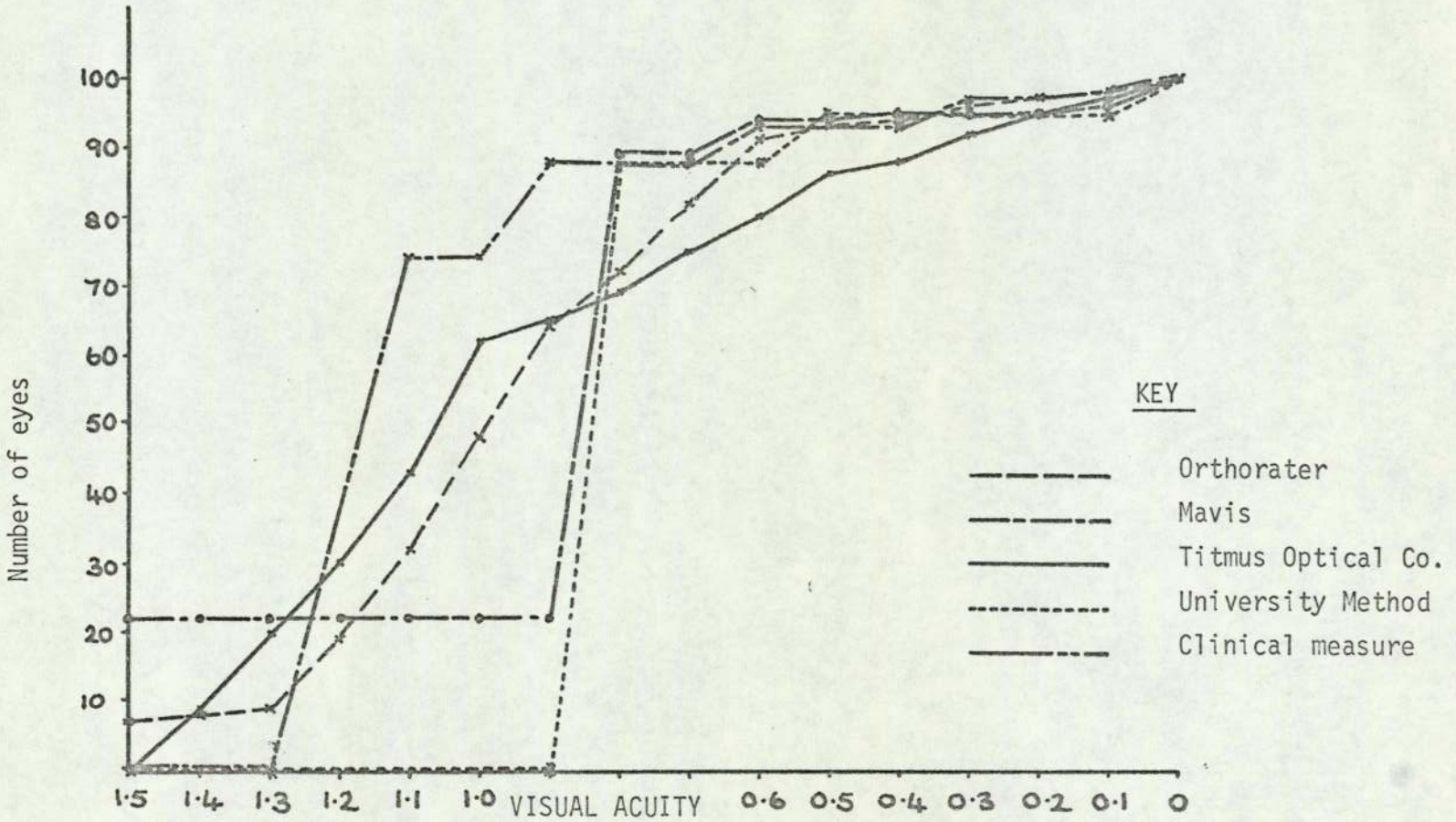


FIG. 8 v Cumulative frequency distributions of visual acuity (left eyes) measured clinically and by each screening method

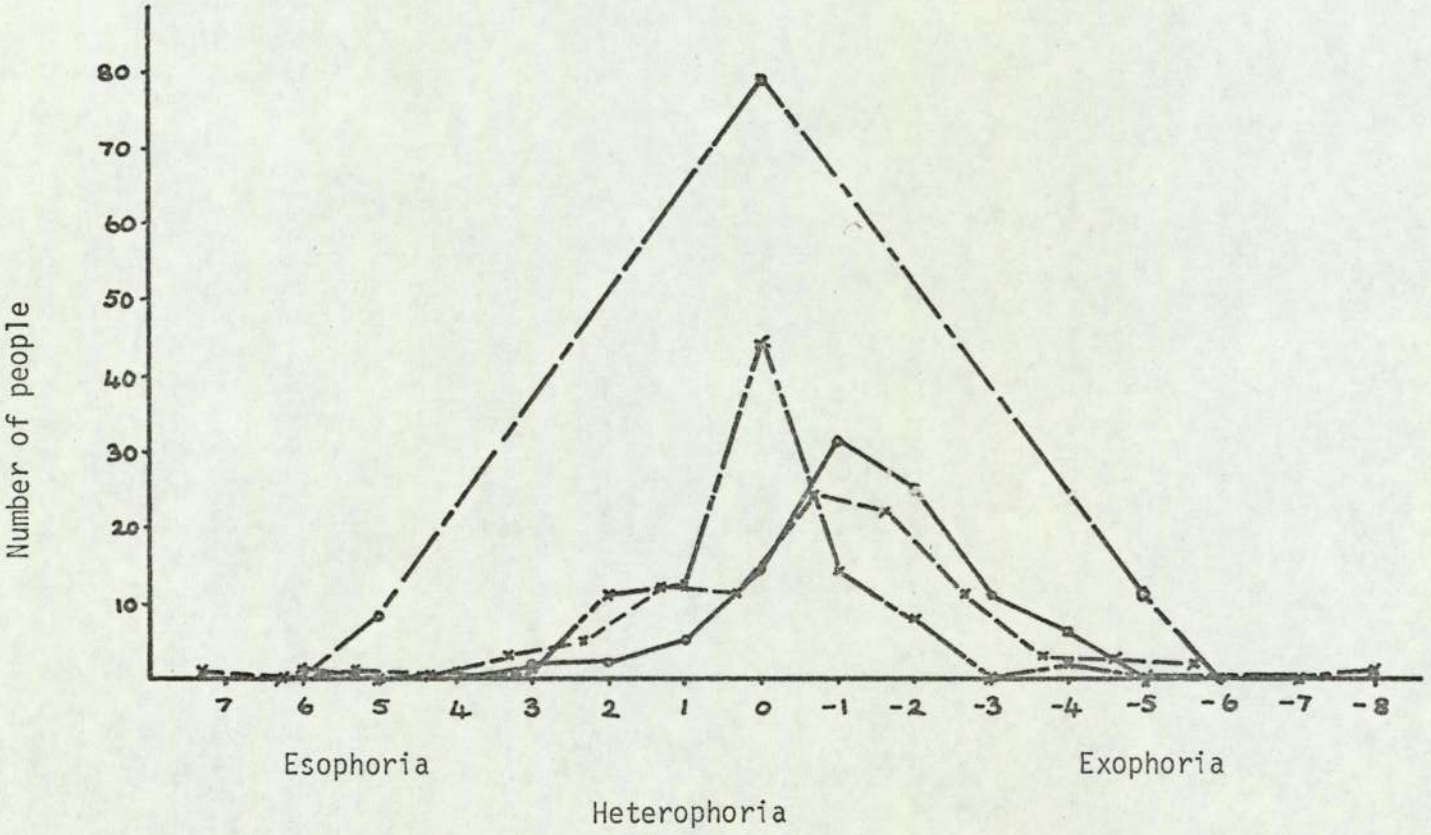


Fig. 8.vi. Frequency distribution of distance, lateral heterophorias

difficulty.

Therefore, the cause of the inaccuracy is probably in the type of target used. An investigation into types of test target was carried out by L. L. Sloane et al (1952). They found that acuity measured with Landolt ring is on the average slightly poorer than that measured by letters, this may account for the discrepancy found in the visual acuity as measured by the Titmus Optical Co. instrument which uses Landolt rings as its test objects. However, Sloane suggests that for use in routine testing the two targets may be considered essentially equivalent. It is interesting to note that Sloane's investigation found that the relationship between acuity measured with letters and with a checkerboard target was not linear, and if ability to recognise complex forms, such as letters, is accepted as a valid measure of visual acuity, then it was concluded that the checkerboard target for some reason overestimates acuities below about the 6/9 level and underestimates those at higher levels. This was not evident in the present experiment because (a) the range of acuities measured was small, and (b) the method of comparison used, the F-test, compares only the mean acuity values, which, if the over and under estimates made by the checker-board target cancel out, will be the same for checker-board and letter acuities.

A second, rather different method of assessing the usefulness of the visual acuity measurements was to compare their value with the amount of ametropia found in the subject by the ophthalmic optician. The comparison was made by specifying the visual acuity as "pass" or "fail", and finding whether the distribution of refractive errors in the "pass" group was significantly different from their distribution in the "fail"

group.

Eight matrices were constructed, for the right and left eyes, for each screener. The following fig. shows the construction of these tables:-

Spherical component of error (dioptres)	V.A. "pass"					V.A. "fail"				
	-ve. cyl. (dioptres)					-ve. cyl. (dioptres)				
	0.0	0.25	0.50	0.75	1.00	0.0	0.25	0.50	0.75	1.00
+										
-										

These matrices were then "collapsed" to determine (a) whether the pass/fail criterion was correlated with the spherical component of refractive error, and (b) whether it was correlated with the cylindrical component of an astigmatic error. Fig.8vi-8xiii show the distribution of passes and fails over the spherical refractive error, for each screener, and each eye separately. Cylindrical errors were treated similarly. χ^2 tests were applied to find whether the distributions were significantly different for the 'pass' group and 'fail' group, on each screener. The following table shows the probabilities associated with the χ^2 values found.

	Sphere		Sphere	
	Right eye	Left eye	Right eye	Left eye
Orthorater	0.2	0.001	0.1	0.01
Mavis	0.02	0.001	0.7	0.01
Titmus Opt.	0.15	$0.01>p>0.001$	0.3	0.01
Univ.Method	0.001	$0.01>p>0.001$	0.25	0.15

FIG. 8 vi

ORTHORATER

Visual Acuity R. eye

Comparison of uncorrected spherical refractive error in the pass group and referred group

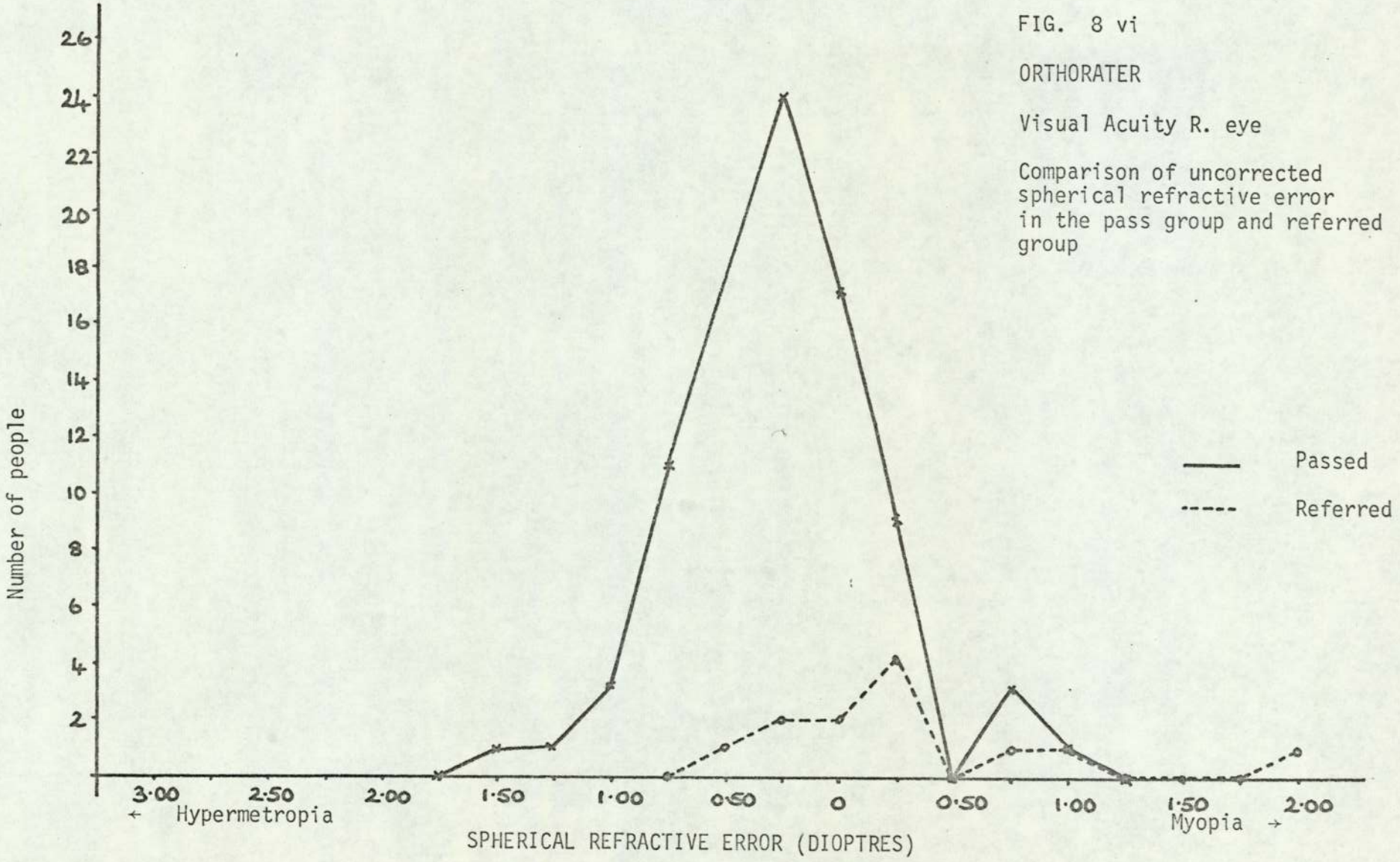


FIG. 8 vii

ORTHORATER

Visual Acuity L. eye

Comparison of uncorrected spherical refractive error in the pass and referred group

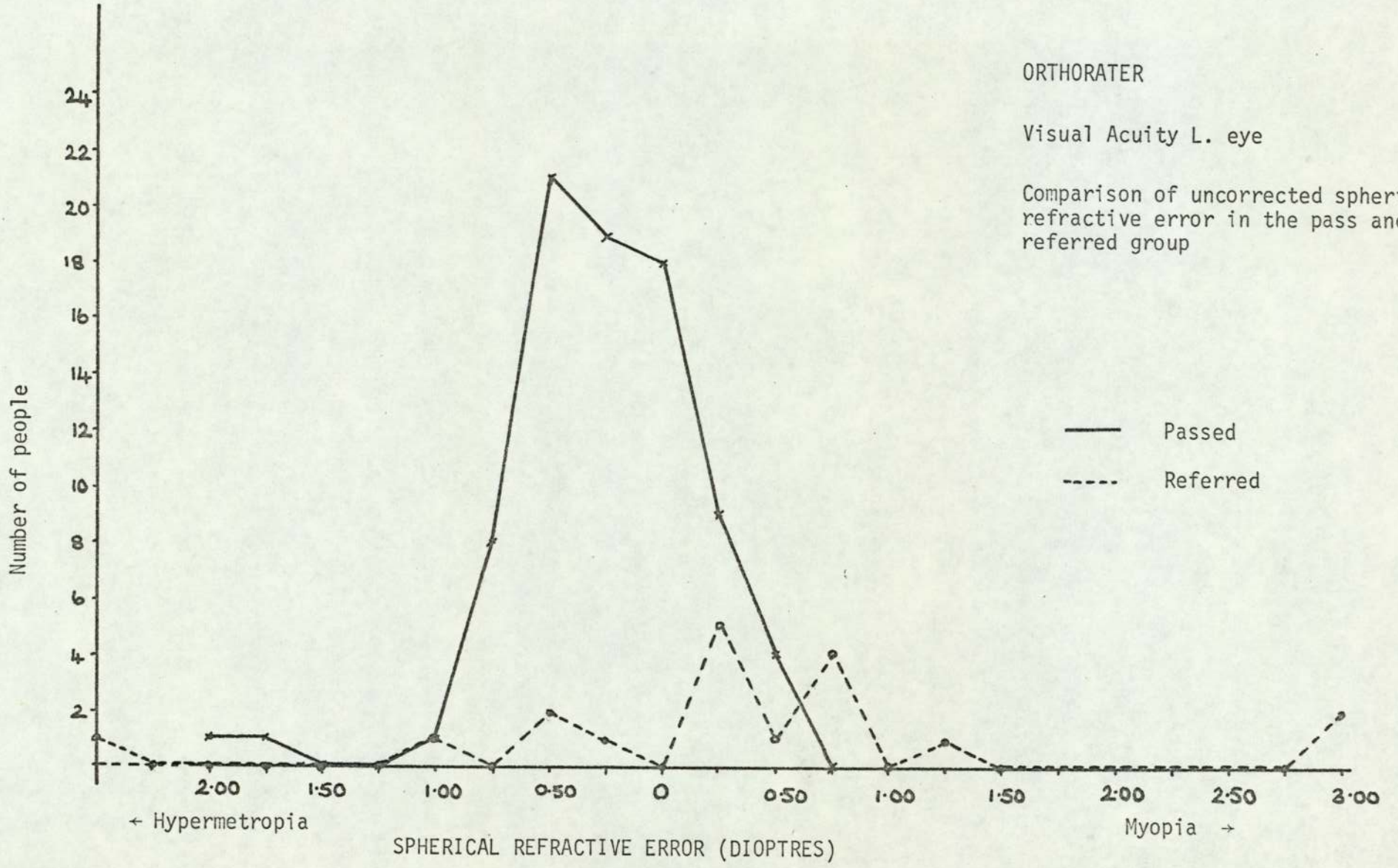


FIG. 8viii

"MAVIS"

Visual Acuity Right Eye

Comparison of uncorrected spherical refractive error in the pass group and the referred group

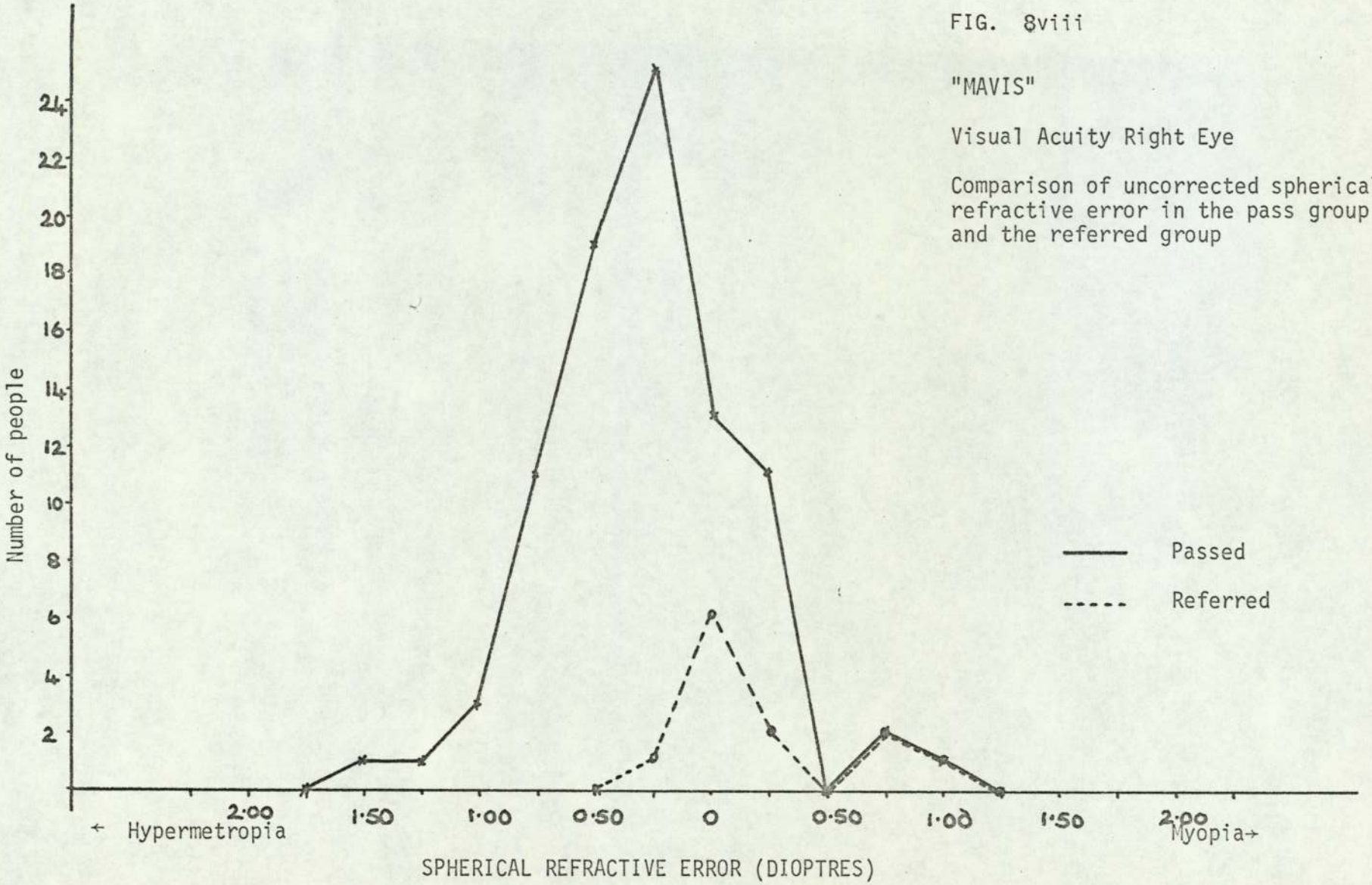


FIG. 8ix

"MAVIS"

Visual Acuity L. Eye

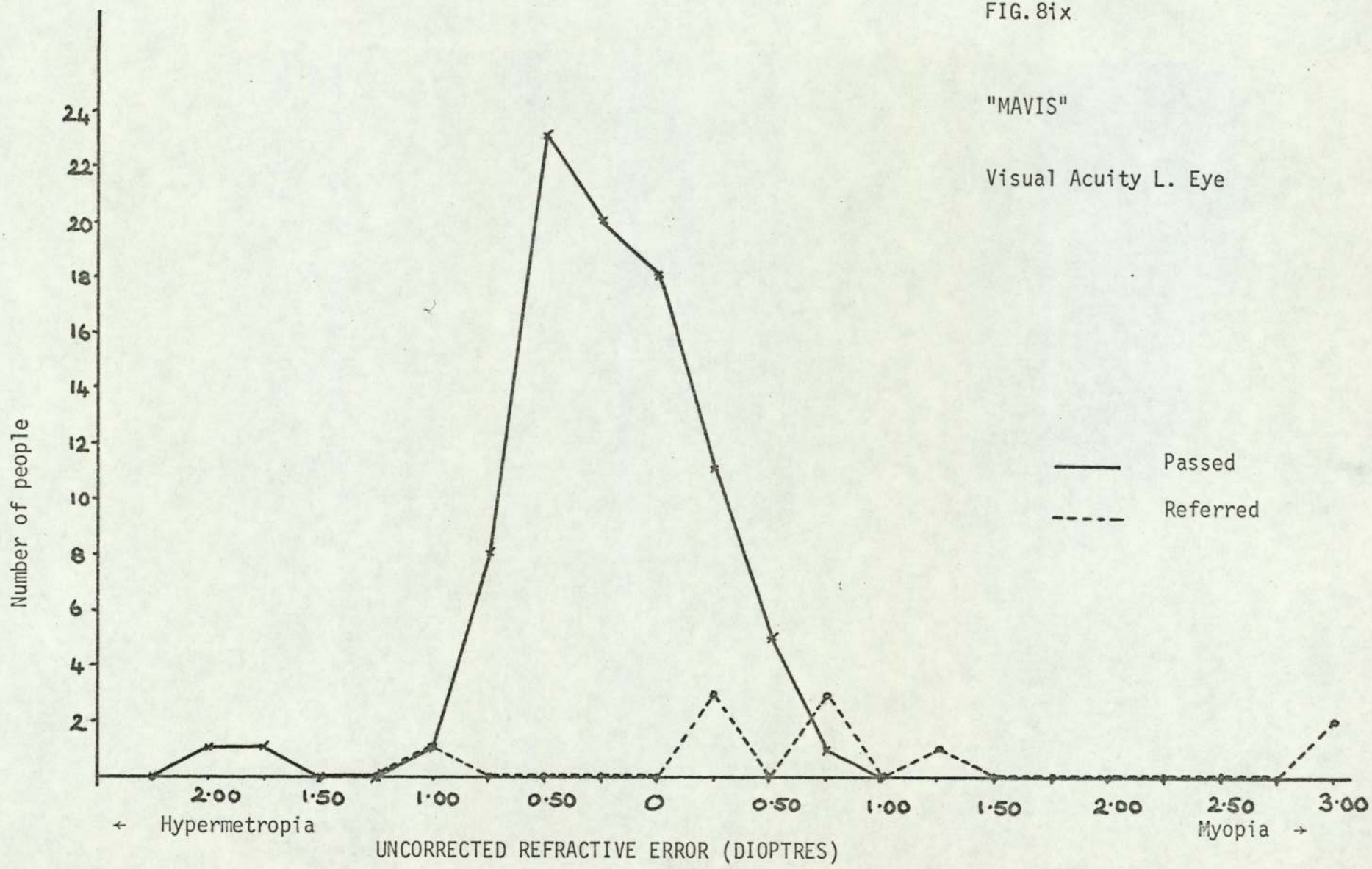


FIG. X

TITMUS OPTICAL CO. SCREENER

Visual Acuity R. eye.

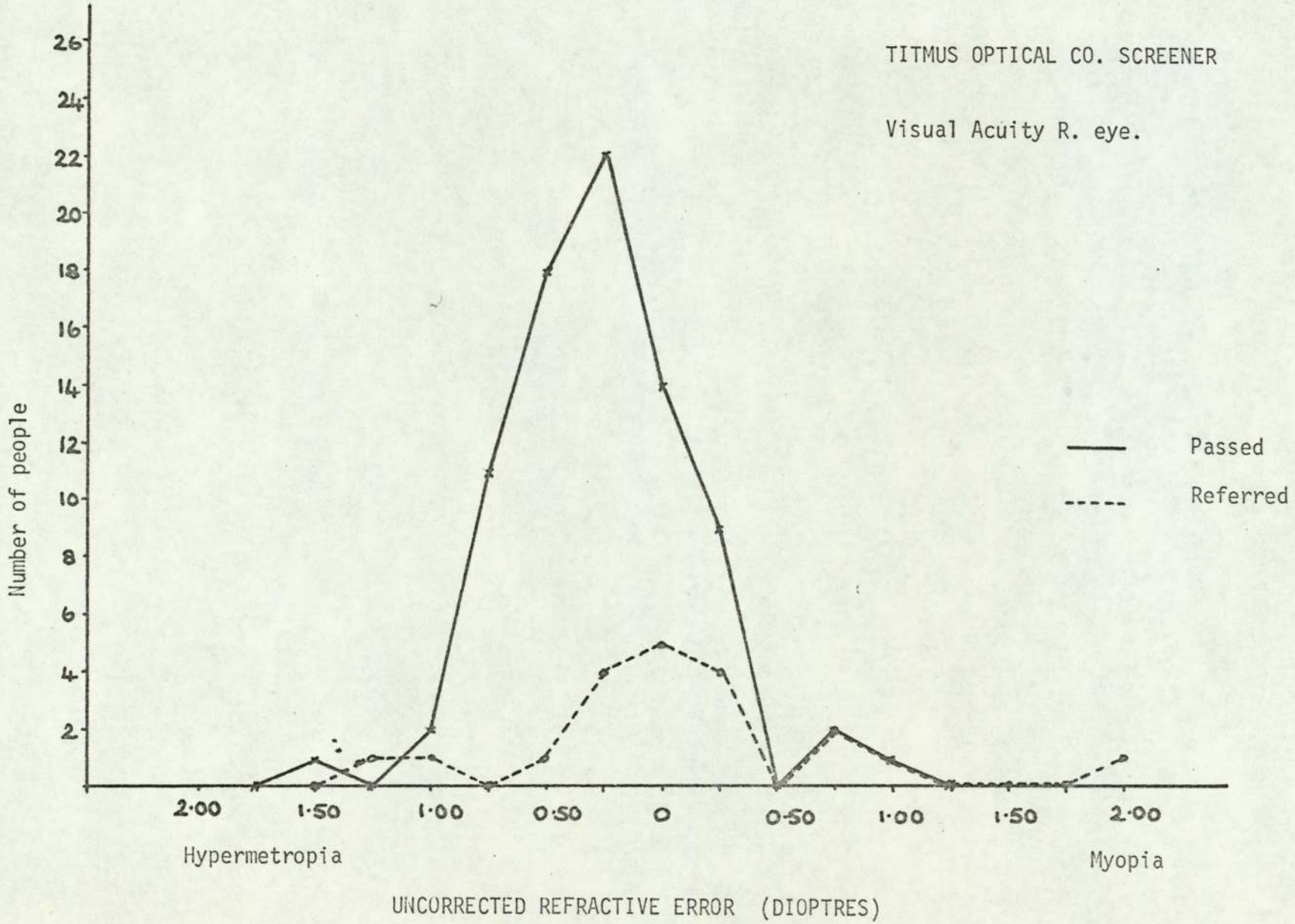


FIG. Xi

TITMUS OPTICAL CO. SCREENER.

Visual Acuity L. Eye

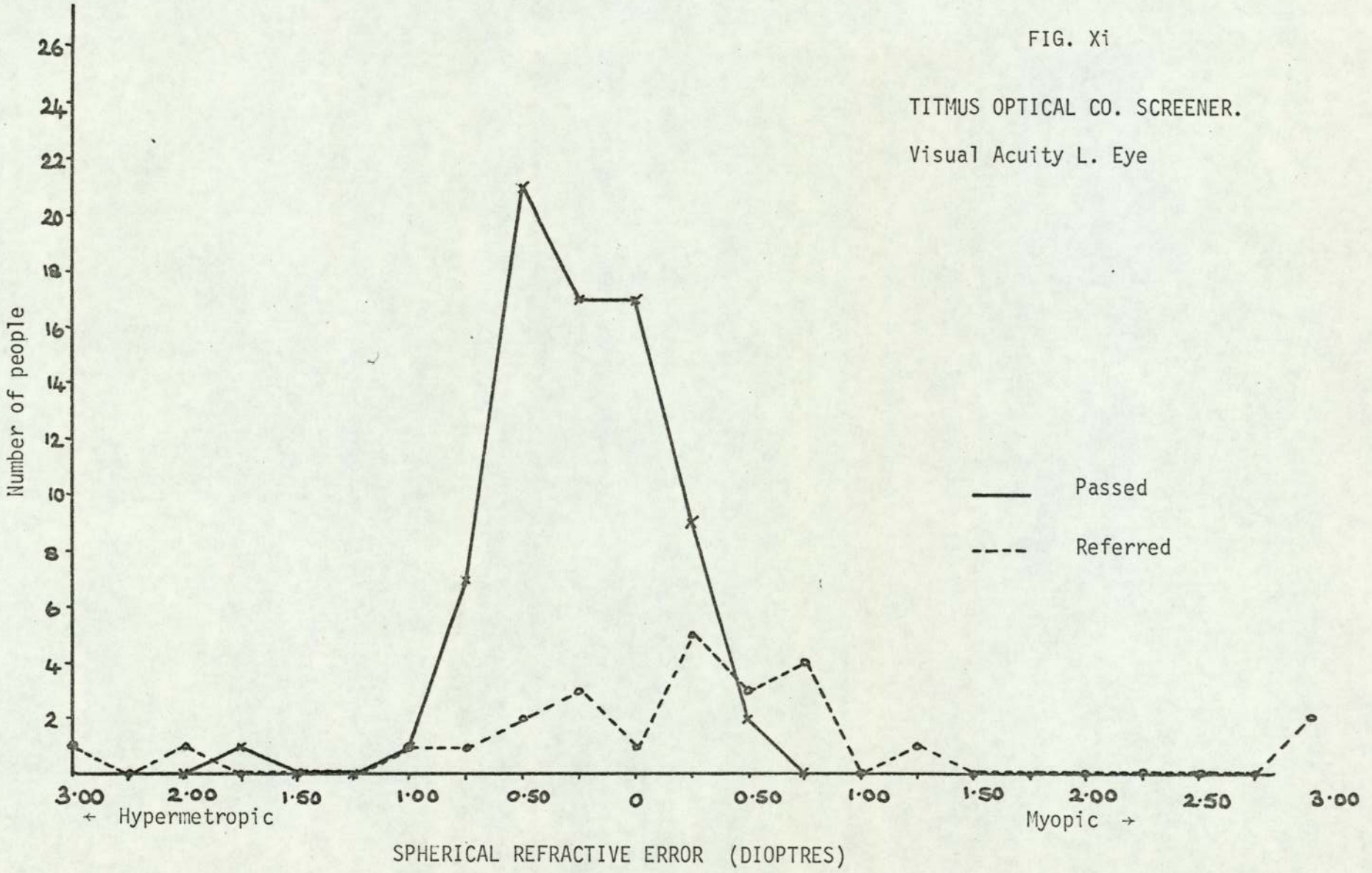


FIG. Xii

UNIVERSITY SCREENING METHOD

Visual Acuity R. eye

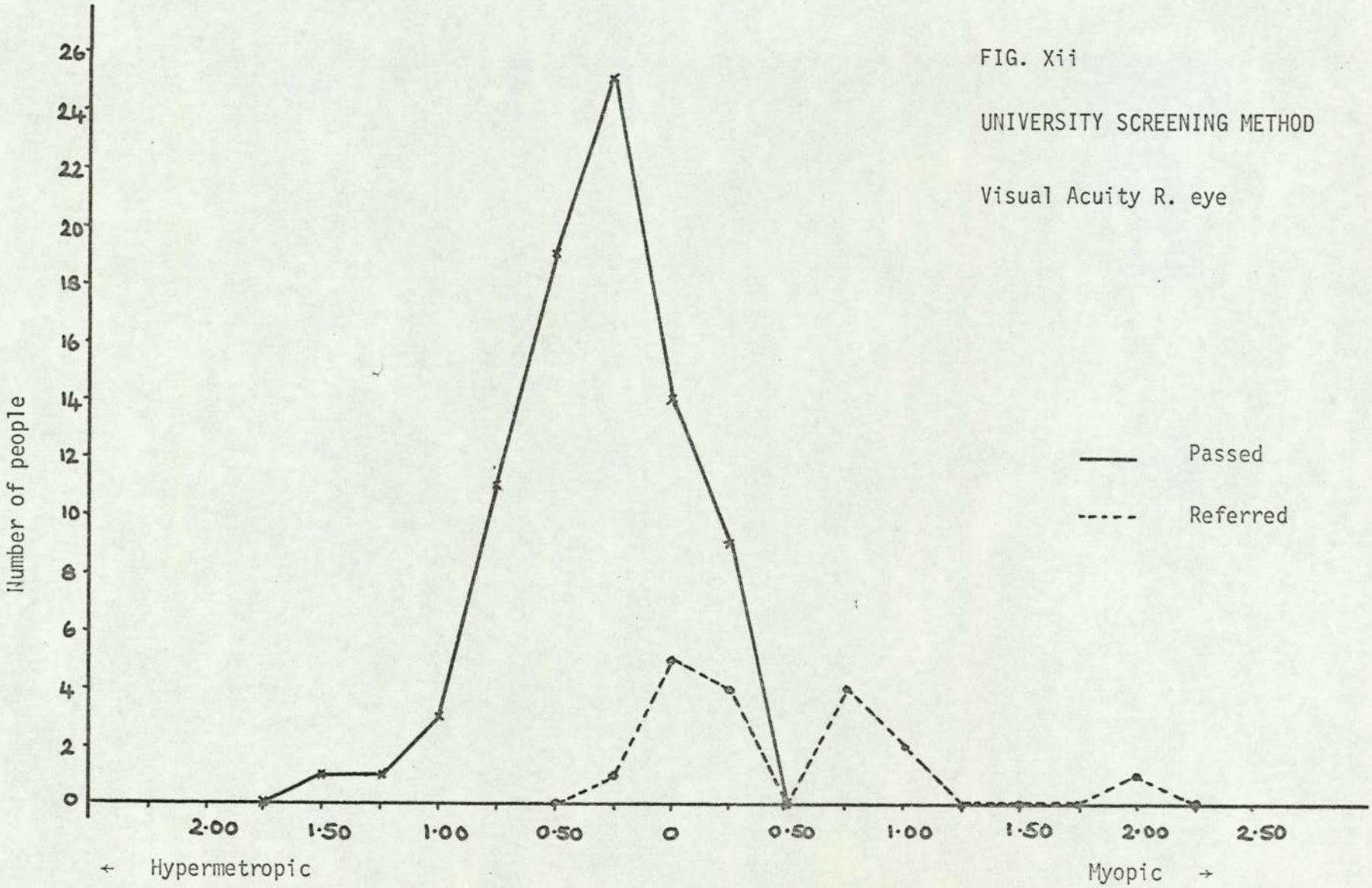
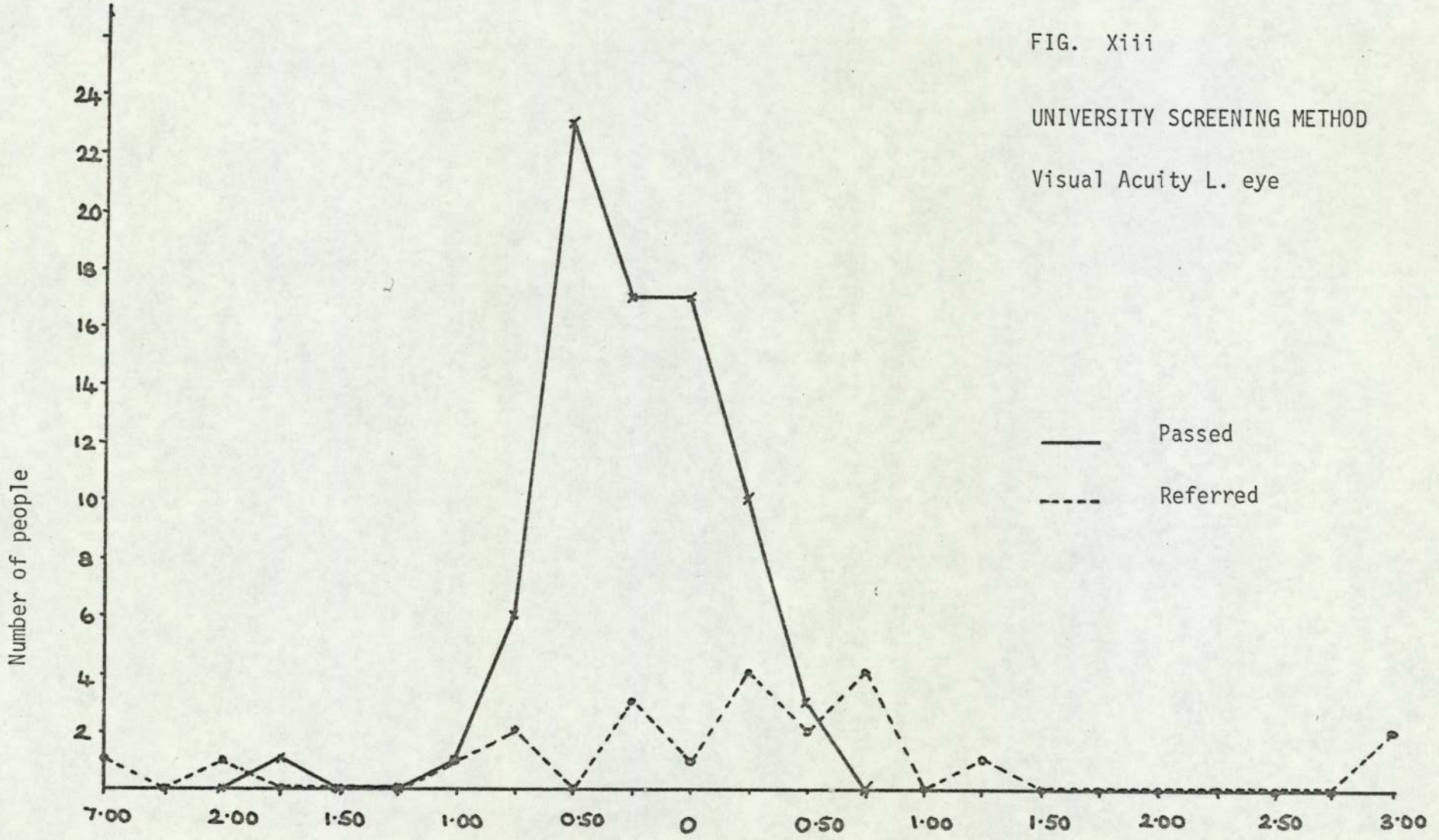


FIG. Xiii

UNIVERSITY SCREENING METHOD

Visual Acuity L. eye



In the right eyes, where spherical errors were found from +1.50 to -2.00 D there is little relationship between the pass/fail criterion and spherical error, except for results obtained from the University technique. However, for the left eye, where the range of errors was larger from +7.00D to -3.00 D, there was a significant difference in the distribution of spherical error for the two groups, "pass" and "fail", for all the vision screening techniques.

When comparing visual acuity results with cylindrical errors no difference was found between the two groups, probably because the maximum amount of cylinder recorded was 1.00D.

Hypermetropia test

The tests for hypermetropia employed by the Mavis and University technique employ a pair of +2.00 D lenses to relax the hypermetrope's accommodation and so permits the test chart to be clearly read. Inability to read the test chart constitutes passing the test. In the case of high hypermetropia, the eye cannot accommodate enough to produce a clear image either in normal viewing or with the help of a +2.00 D lens, and so will appear to pass this test. This type of error will be found in the visual acuity tests; however, this limitation of the +2.00 D test must be remembered.

The "pass" group was compared with the "fail" group for this test in terms of the distribution of refractive errors within the two groups, in the same way as the visual acuity tests. It will be expected that where there are no very high hypermetropes, then the difference in distributions between the two groups will be significant; however, the presence of high hypermetropia will weight the amount of

hypermetropia in the "pass" group making the difference between the two groups less significant (or even significant in the opposite direction, more hypermetropes passing than failing).

The results for the right eyes, where the incidence of high hypermetropia was low, show a significant difference between the "pass" and "fail" groups. However, as expected. the results for the left eyes, where the highest hypermetrope is +7.00D, show no significant difference between the distributions of refractive errors over the two groups.

Table 8i shows the probabilities derived from the χ^2 values obtained from the comparison of the pass and fail groups for this test with respect to refractive errors.

	Spherical error		Cylindrical error	
	Right eye	Left eye	Right eye	Left eye
Mavis	0.001	0.99	0.75	0.85
University technique	0.001	0.05	0.15	0.6

Table 8 i.

Distance lateral heterophoria

This measurement was made by all the screening methods except the University technique. The F-test comparing the mean heterophoria value for each screening method and for the clinical method, found that the null hypothesis that "there was no significant difference between means" could be rejected. The t-test, applied to compare the clinical lateral heterophoria values with each screener heterophoria value, found that the Titmus Optical Co. screener had a mean heterophoria value significantly different from that measured clinically. However, for each of the other screening methods the difference was not found to be significant, and the null hypothesis was not rejected

for them.

The Titmus Optical Co. instrument showed a bias towards exophoria. The graph showing the heterophoria distributions for each screener shows this tendency. (fig. 8vi. P80.)

Near lateral heterophoria

This measurement was made by all the screening methods. An F-test applied to compare the mean heterophoria values found by each screening method rejected the null hypothesis that "there is no significant difference between means". T-tests to compare pairs of means showed that both the Mavis and Titmus Optical Co. Screener had results significantly different from the clinical results. The error may be due to proximal convergence induced by using enclosed boxes for the measurements. The charts in both the Orthorater and Mavis are decentred to allow for this effect, but Cutler and Davey in 1965 found that the Mavis was not decentred enough, and found a shift towards esophoria in the measurements; from this investigation it seems probable that the Titmus Optical Co. Screener also has the same fault.

Distant vertical heterophoria

The distance vertical heterophoria measures were made by all but the University screening technique. The F-test rejected the null hypothesis that there was "no significant difference between means". More detailed examination using the t-test showed that all the screening means were significantly different from the clinical value - the Orthorater showing a shift towards right hyperphoria and the Mavis and Titmus Optical Co. screener show a shift towards left hyperphoria. When studying the Mavis, Cutler and Davey (1965) also found a shift towards left hyperphoria. These erroneous measures

are probably due to the design of the tests which leave the instructions open to individual interpretation by the subject.

Near vertical phorias

The F-test applied to the only three measurements made of near vertical phoria, the Orthorater, Mavis and clinical values, found a significant difference between means. The t-test showed this to be due to an error in the Mavis which produced results showing a considerable shift towards left hyperphoria probably due to the same design fault as in the distance measurement.

Colour vision tests

To find the validity of colour vision tests a value for ϕ was found for each screener when compared with the clinical Ishihara results. The values found were as follows:-

		Sensitivity	Specificity
Orthorater	ϕ = 1.00	100%	100%
Titmus Opt.Co.	ϕ = 0.5	37.5%	98.6%
University technique	ϕ = 0.9	100%	97.8%

The Mavis does not test colour vision. It was found that the Orthorater showed perfect agreement with the clinical Ishihara findings, and was 100% sensitive and specific.

Conversely the Titmus Optical Company screener was not very sensitive, that is its ability to fail a colour defective was poor, although its ability to find a non-colour defective was high. It did not fail any subject with normal colour vision. In order to fail this test

a subject must mis-read more than three of the eight test cards. If this number is reduced, the test is made more difficult to pass, then there is a marked change in sensitivity and specificity, e.g.

If any card read incorrectly constitutes failure of the test, then SENSITIVITY = 100%

SPECIFICITY = 93.5%

This is a considerable improvement. The best compromise is probably two cards mis-read constituting failure of the test. This gives:

SENSITIVITY = 87.5%

SPECIFICITY = 95.6%

The University screening technique was 100% sensitive but tended to over-refer and so had a lowered specificity of 97.8%.

Binocular visual acuity

This was measured only by the Orthorater and Titmus Optical Company instruments, and was not recorded clinically. Therefore, no estimate could be made of their accuracy. However, the two screener results were compared with each other, to find out whether they measured essentially the same thing. As the scales used for the measurement were similar a t-test was applied to the acuity distributions obtained from the instruments. It was found that the distribution means were not significantly different from each other.

Near visual acuity and stereopsis

Near visual acuity was measured by all the screening

methods, but was not recorded clinically.

Stereopsis was measured by the Orthorater, Titmus Optical Company instrument and the"Mavis", but again was not recorded clinically.

The unstandardised scaling makes accurate comparison of these measurements impossible.

8.iii. REFERRAL STANDARDS

There have been many attempts to standardise the cut-off points which determine whether a subject is referred by a particular screening test. These have mostly been in the form of questionnaires set out to obtain the opinions of various members of the optical profession. All of these questionnaires have found considerable difference of opinion as to the placing of the cut-off point. When examining patients individually a picture of what is "normal" for that patient is built up. This picture may be very different from that of another patient, it being based on the patient's age, occupation, general health etc. It is this difference between individuals which makes mass assessments of any kind very difficult and inaccurate. Crane in 1950 stated that: "One of the difficulties in evaluating vision testing procedures has always been that no definitions have been established of the smallest degree of abnormality of the eye for which observation or treatment is needed". In my opinion no definition of this sort can be made because people are so varied in their individual tolerance of abnormality. Instead of obtaining the cut-off points by a concensus of professional opinion, a different set of criteria should be applied. These criteria depend upon the sensitivity and specificity of a set of screening tests. These sensitivity and specificity values are laid down in a certain pattern, the pattern is influenced by the choice of tests and their accuracy and the prevalence of the condition being tested among the population being screened. Hence for a standard set of tests on a specific population the pattern is constant. Any alteration of cut-off points changes

the sensitivity and specificity values only within the established pattern. Hence it can only alter the balance of false to true referrals and non-referrals. It should be upon this balance that decisions regarding the cut-off points should be made. The effect of false test results upon the community must be weighed against the value of finding true positives and negatives.

Blumberg (1957) listed some items which should help in this evaluation. He sites the general case of screening for disease.

(a) What is the outlook for a person with the disease? The value of a true positive increases as the patient's chances for cure or shorter convalescence are improved by early detection. In fact, when no health benefits are attributable to the finding of cases, the advisability of screening is doubtful.

(b) What facilities exist for treating cases found?. The value of a true positive is reduced if inadequate facilities exist for treating those found. As far as direct benefits go, it is only profitable to find as many cases as may be treated. (In the long run, finding larger numbers of cases may lead to the provision of more adequate treatment facilities, which in turn may help the cases already found, provided they live that long).

(c) What mental state accompanies knowledge or suspicion of the disease? If suspicion of a disease is accompanied by considerable anxiety that may in turn be debilitating, then demonstration of a true negative could provide valuable reassurance and health benefits. Ordinarily, however, very little good (or harm) is done by finding a true negative result. On the other hand, the fear from being falsely considered positive might very well cause

undue anxiety and thus be of direct harm to those who have been placed in this category, but if the notice which directs positives to seek further diagnostic study is tactfully and intelligently written, this danger from false positives can be minimised.

(d) Who is going to do the diagnostic follow-up? False positives burden the diagnostic facilities. If the facilities are adequate and the diagnostic studies quite inexpensive, then false positives may be less harmful. False positives also serve to discredit screening procedures and screeners in the eyes of those screened and medical practitioners. In a closed community, such as the armed forces, where follow-ups are done entirely at the expense of the community and not of the individual, false positives may be less harmful.

(e) What is the likelihood of repeat screening within the community? If it is unlikely that repeat screening will be carried on within a short period of time, then false negatives could be extremely detrimental. On the other hand, if screening will be repeated in a short period and the disease is not communicable or rapidly progressing, then false negatives would not be so harmful, since there may be a fair likelihood of uncovering the disease the next time.

(f) Are healthy individuals being sought? Sometimes screening procedures are adopted to find healthy rather than sick individuals. This may be the case in selecting people for certain jobs, or for the armed services, as well as in screening life insurance applicants. In these cases false negatives are very costly whilst false positives may not be.

The six questions posed above by Blumberg will now be applied to the visual screening of freshmen at the University of Birmingham:

(a) What is the outlook for a person with the disease? If a person is found to be ametropic then this can be corrected by an optical appliance. Faults of binocular vision can be treated, but probably not very successfully in people of this age group. If the ametropia is uncorrected the symptoms will be blurred vision with possible asthenopia. If the subject has a pathological condition then early treatment is the most effective. If the disease is missed then the prognosis could be poor, even fatal in rare cases.

(b) What facilities exist for treating cases found? Within the University campus there is a Health Centre which provides comprehensive treatment facilities both for out-patients and in-patients. These include an ophthalmic clinic attended by ophthalmic opticians and an ophthalmic medical practitioner. Outside the campus the National Health Service provides the Supplementary Ophthalmic Services, and the Hospital Eye Service for more urgent cases. Therefore, any person found to require a follow-up examination and treatment could get it fairly promptly.

(c) What mental state accompanies knowledge or suspicion of the disease? Generally people tend to be concerned about their eyes and their sight. Any suspicion that something may be wrong with either makes them anxious.

(d) Who will do the follow-up examination? Follow-ups will generally be done at the expense of the community, with possibly some expense to the individual. False positives would not be popular either with the person carrying out the examination, or with the student himself.

(e) What is the likelihood of repeated screening? Unless

the student is screened before taking a job, it is unlikely that he will ever have a visual screening examination again. Therefore, this may be the last opportunity for finding visual disorders. Hence the false passes should be kept to a minimum.

Before making any decisions about the positioning of the cut-off points, the sensitivity-specificity pattern of the existing screening should be found, and the effects on this, of varying the cut-off points of the sub-tests.

THE SENSITIVITY-SPECIFICITY PATTERN OF FOUR SCREENING METHODS

Using the results of the experiment to find the efficiency of the screening techniques described previously (P 49) it was possible to study the effect of changing the cut-off points of the sub-tests on the sensitivity and specificity values of each of them.

Changes in the prevalence of an abnormality can change the sensitivity-specificity pattern. The experimental sample enables the prevalence to be kept constant.

The screening and clinical examination results for each subject were coded on to edge-punched cards. This facilitated the re-sorting necessary to produce the four groups (true and false referrals, and true and false non-referrals) whose magnitudes changed as the cut-off points were moved.

ORTHORATER

SUB-TEST	NATURE OF CUT-OFF POINT VARIATION	OVERALL SENSITIVITY AND SPECIFICITY	
		Sensitivity	Specificity
	With recommended cut-off points	54.8%	65.2%
DISTANCE VISUAL ACUITY	Made more difficult: V.A.less than 6/6 fails	93.5%	28.1%
	Made easier: V.A.less than 6/12 fails	45.2%	73.9%
DISTANCE VERTICAL HETEROPHORIA	Test omitted	54.8%	65.2%
DISTANCE LATERAL HETEROPHORIA	Made difficult: Fail more than: 2Δ esophoria 2Δ exophoria	64.5%	53.6%
	Test omitted	54.8%	65.2%
NEAR VISUAL ACUITY	Test made more difficult: V.A.less than 14/14 equivalent fails	80.6%	43.5%
	Test made easier: V.A. less than 14/24 equivalent fails	51.6%	66.7%
NEAR VERTICAL HETEROPHORIA	Test omitted	54.8%	65.2%
NEAR LATERAL HETEROPHORIA	Test made more difficult: Fail more than: 2Δ esophoria 2Δ exophoria	64.5%	37.7%

TITMUS OPTICAL COMPANY SCREENER

SUB-TEST	NATURE OF CUT-OFF POINT VARIATION	OVERALL SENSITIVITY AND SPECIFICITY	
		Sensitivity	Specificity
	With recommended cut-off points	71.0%	62.3%
DISTANCE VISUAL ACUITY	Test made more difficult: V.A. less than 6/6 fails	80.6%	46.4%
	Test made less difficult: V.A. less than 6/12 R or L or 6/10-5 both eyes, fails	45.2%	68.1%
DISTANCE STEREOPSIS	Test made more difficult: Cut-off between 70 seconds and 50 seconds of arc, angle of stereopsis	71.0%	56.5%
DISTANCE VERTICAL HETEROPHORIA	Test excluded	71.0%	62.3%
DISTANCE LATERAL HETEROPHORIA	Made more difficult: Fail more than: 3Δ esophoria 3Δ exophoria	71.0%	60.9%
	Test excluded	71.0%	62.3%
NEAR VISUAL ACUITY	Made more difficult: Less than 14/14 (Snellen equivalent) fails	74.2%	57.9%
	Made less difficult	67.7%	68.1%
NEAR LATERAL HETEROPHORIA	Made more difficult Fail more than: 3Δ esophoria $4\frac{1}{2}$Δ exophoria	74.2%	59.4%
	Test excluded	71.0%	65.2%

MASTER VISION SCREENER

SUB-TEST	NATURE OF CUT-OFF POINT VARIATION	OVERALL SENSITIVITY AND SPECIFICITY	
		Sensitivity	Specificity
	With recommended cut-off points	48.4%	65.2%
DISTANCE VISUAL ACUITY	Made more difficult: Fail 6/7.5 or less	100%	5.8%
	Made less difficult: Fail 6/14 or less	41.9%	66.7%
HYPERMETROPIA	More difficult: Fail +1.00	48.1%	42.0%
	Test omitted	48.4%	66.7%
DISTANCE VERTICAL PHORIA	Test omitted	48.4%	65.2%
DISTANCE LATERAL PHORIA	Made more difficult: Pass orthophoric only	54.8%	53.7%
	Test omitted	48.4%	65.2%
NEAR VISUAL ACUITY	Made more difficult: Fail less than N.4.	58.1%	
	Test omitted	48.4%	66.7%
NEAR VERTICAL PHORIA	Test omitted	45.2%	73.9%
NEAR LATERAL PHORIA	Made more difficult: Pass orthophoric only	61.3%	52.2%
	Test omitted	45.2%	73.9%
BINOC. V.A.and SIMULTANEOUS FOVEAL VISION	Test omitted	48.4%	65.2%

UNIVERSITY SCREENING METHOD

SUB-TEST	NATURE OF CUT-OFF POINT VARIATION	OVERALL SENSITIVITY AND SPECIFICITY %	
		Sensitivity	Specificity
	With recommended cut-off points	71·00	82·6
HYPERMETROPIA TEST	More difficult: Reading either 6/12 letter = Fail	74·2	79·7
	Test omitted	71·0	82·6
DIST. V.A. (T.I.B)	More difficult: Any letter read wrongly = Fail	71·0	75·4
3^{Δ} EXTENSION OF T.I.B. TEST	Test extension omitted	71·0	79·7
NEAR TEST	Test omitted	71·0	84·0
CONVERGIOMETER	Fail $>1^{\Delta}$ esophoria or exophoria	80·6	49·3
	Test omitted	67·7	84·0

CONCLUSIONS

It was found that certain tests could be entirely omitted from the complete test batteries, without affecting the overall sensitivity and specificity of the test.

E.g. SCREENER	SUB-TEST
Orthorater	Near vertical phoria test
	Dist. vertical phoria test
	Dist. lateral phoria test
Titmus Optical Co. Screener	Dist. Lateral phoria test
	Dist. Vertical phoria test
"Mavis"	Dist. Vertical phoria test
	Dist. Lateral phoria test
	Binoc. V.A. and simultaneous
	foveal vision test

Conversely, slight changes of cut-off point in other sub-tests produced drastic changes of sensitivity and specificity:

E.g. SCREENER	SUB-TEST
Orthorater	Distance V.A. Test
	Near V.A. Test
Titmus Optical Co. Screener	Distance V.A. Test
	Near V.A. Test
"Mavis"	Distance V.A. Test
	Hypermetropia Test
	Near V.A. Test
	Near Lateral Phoria Test

This supports the theory that these are the major tests and demonstrates that some tests actually waste time, and contribute nothing to the overall effectiveness of the screening tests, when testing this particular population.

9. CONCLUSIONS AND SUGGESTIONS FOR FUTURE RESEARCH

The two main objectives of this investigation, as stated in section 3, were:-

i. To analyse the screening results at the University of Birmingham for the years 1958 - 67.

ii. To find whether the screening programme used was the one most suited to the situation. If not, to make modifications.

In trying to fulfil these objectives the University screening technique was compared with other, accepted, techniques, and with a full clinical examination. From these investigations the screening technique emerged remarkably well, and the proprietory screeners were shown to be somewhat inefficient and, therefore, it is to them that any modifications should be made.

No screening method is perfectly efficient, even when specifically designed to screen a particular section of the population. Experiment has shown that this lack of efficiency is not due to inaccuracy of the sub-tests, or poor repeatability, for these aspects of the tests seem fairly reliable. Any errors found were too small to produce the gross inefficiency demonstrated. The reasons for the inefficiency must lie elsewhere:-

(a) Number of sub-tests

The first screening test used was simple measurement of distance visual acuity with a Snellen chart. This was found to be inadequate, so other tests were added to compensate for the inadequacy. This trend, however, has led screening methods to use a large number of sub-tests, producing unnecessary complexity, wasting

time and detracting from the overall efficiency of the screener. These redundant tests should be omitted, leaving the nucleus of important tests, which can then be used selectively, bearing in mind the population being screened.

(b) Technician training

The greater efficiency of the University technique is probably due to there being a skilled examiner present to make decisions on borderline cases. His presence, however, is contrary to the screening precept that a non-skilled person should apply the tests, to release the practitioner for work which is more demanding of his skills. There is a great diversity of opinion as to how much training a technician should be given. Basically the amount of training required depends on the person's aptitude for the work, and the type of population he will be examining.

Crane (1952), when discussing the use of the Telebinocular in screening school children in the United States, says that for testing the older, sixth-grade children successfully it was not necessary for the tester to have elaborate training or experience. However, with young first-grade children the testing can be done only by a tester who has had considerable experience in such methods. Generally speaking, young children and elderly people are the most difficult to screen, and yet they will derive the most benefit from resular testing. The problem of training testers should be investigated by someone qualified to carry out work of this nature.

(c) Instructions

These should be made unambiguous and clear, leaving little room for individual subjective interpretation by either the subject or examiner.

ESTABLISHING THE SENSITIVITY, SPECIFICITY PATTERN

For any given screening programme there are optimal levels of sensitivity and specificity, decided by consideration of Blumberg's criteria (see page 98). Therefore the sensitivity and specificity levels obtained by the screening technique should be measured and adjusted by altering the cut-off points of the sub-tests to conform as nearly as possible to the optimal levels. The prevalence of visual defects in the population effects the sensitivity and specificity levels; therefore any changes in prevalence should be noted and adjustments made to the cut-off points to restore the desired sensitivity and specificity levels.

When analysing the screening results at the University of Birmingham it was found that the prevalence of defects was changing (assuming that the screening itself remained constant). A regular check should be made of the prevalence levels in order that the optimal sensitivity and specificity values may be maintained. It is unfortunate that the most convenient check which can be made is by using the screening test itself. It would be more accurate to carry out a clinical examination on a random sample of students, to see whether, for example, there has been an increase in the incidence of myopia, heterophoria etc. It is important to know the sensitivity and specificity pattern which can be expected for any screening procedure. Hence before it is released for general use a comparative study should be made to find this pattern. The University of Birmingham screening method is an example of a technique which has been used for many years with the knowledge of its reliability and under-referral rate, and hence no knowledge of its efficiency, sensitivity or specificity.

FOLLOW-UP EXAMINATIONS

The greatest cost in a screening programme is not in the screening procedure itself but in the follow-up examinations, which are an essential part of the programme. This fact is often forgotten, as its cost is met normally by the state or by the individual who has been referred. It is most important that a referred subject receives the follow-up clinical examination. This is affected by the necessity to give complete freedom of choice to the individual as to whether he undergoes a follow-up examination. It was disturbing to find, via a questionnaire, that many of the students referred by the screening did not present themselves for the follow-up examination. However hard one strives to produce a reliable screening technique, if the subject does not follow the advice of the findings, then it is useless. The answer to this problem is to ensure that the choice made by the subject is based on a full knowledge of the reliability of the screening test as a guide. It seems that too much emphasis is placed on saying that the screening is only a crude test, and not enough emphasis on persuading the subject that a screening referral is worth following up.

FREQUENCY OF SCREENING

As screening is not perfect, it should be applied to the population more than once or twice in their lives, so that an "under-referral" who is missed the first time may eventually be discovered.

APPENDIX I

DESCRIPTIONS OF THE THREE MANUFACTURED VISUAL SCREENERS

Full descriptions of the three proprietary visual screeners can be found in the manufacturers' manuals.

ORTHORATER

The model used throughout these experiments was the MODIFIED ORTHO-RATER, manufactured by Bausch and Lomb.

The instrument is manually operated, the slides for each sub-test being placed into position by the tester. The slides are transilluminated.

There are twelve sub-tests: visual acuity at distance and near for the right and left eyes separately and together - the chart used is the "checkerboard", which is claimed to be the purest measure of retinal resolution; vertical and lateral muscle balance at both distance and near; a distance test of depth perception; and a colour vision test, consisting of photographically reproduced Ishihara colour plates.

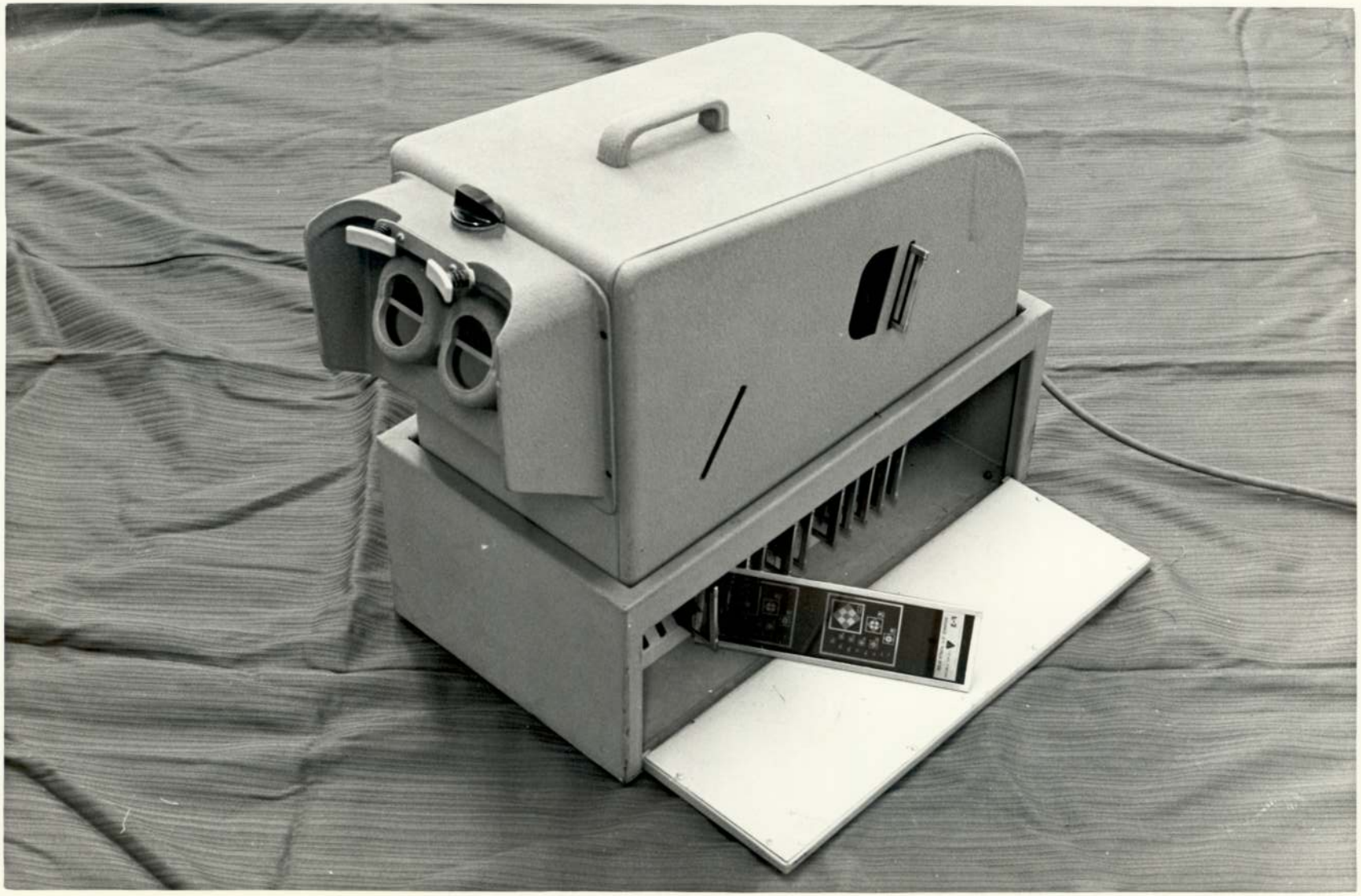
Profiles of cut-off points recommended for occupational requirements have been made.

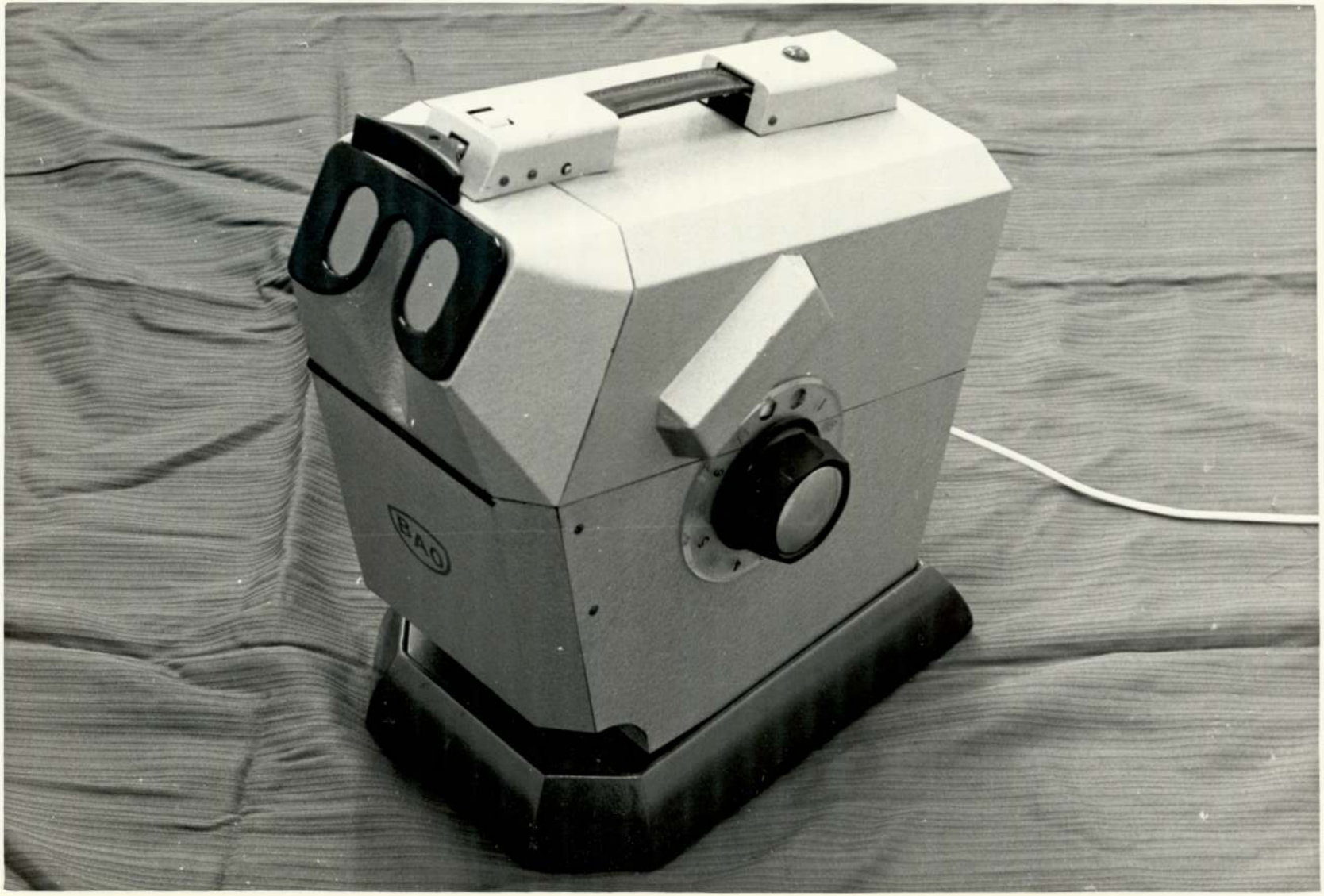
MASTER VISION SCREENER

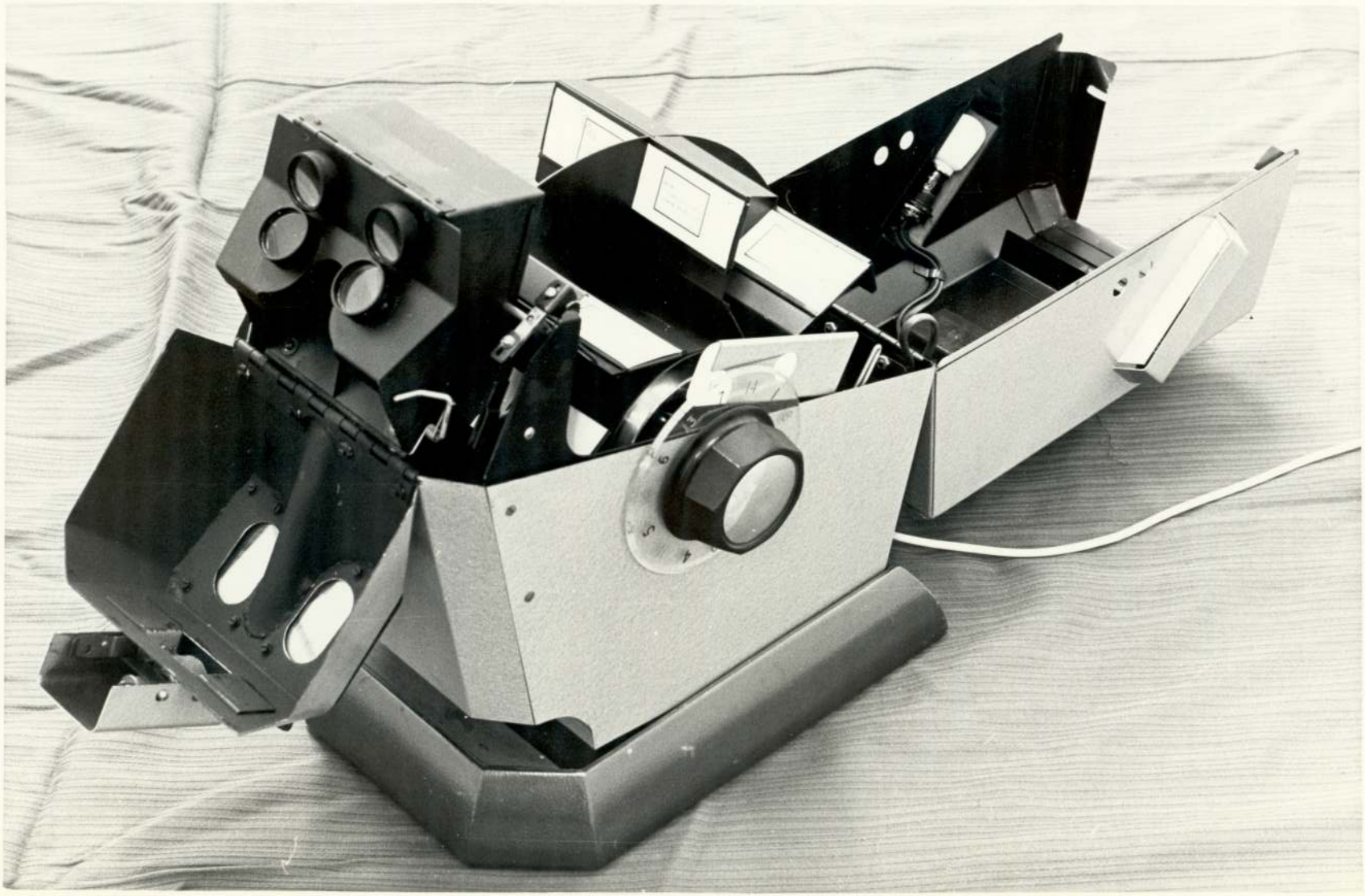
Designed by R. J. Fletcher in conjunction with J. & R. Fleming Ltd. this instrument consists of a manually rotated drum which bears the test targets. The targets are illuminated from the side.

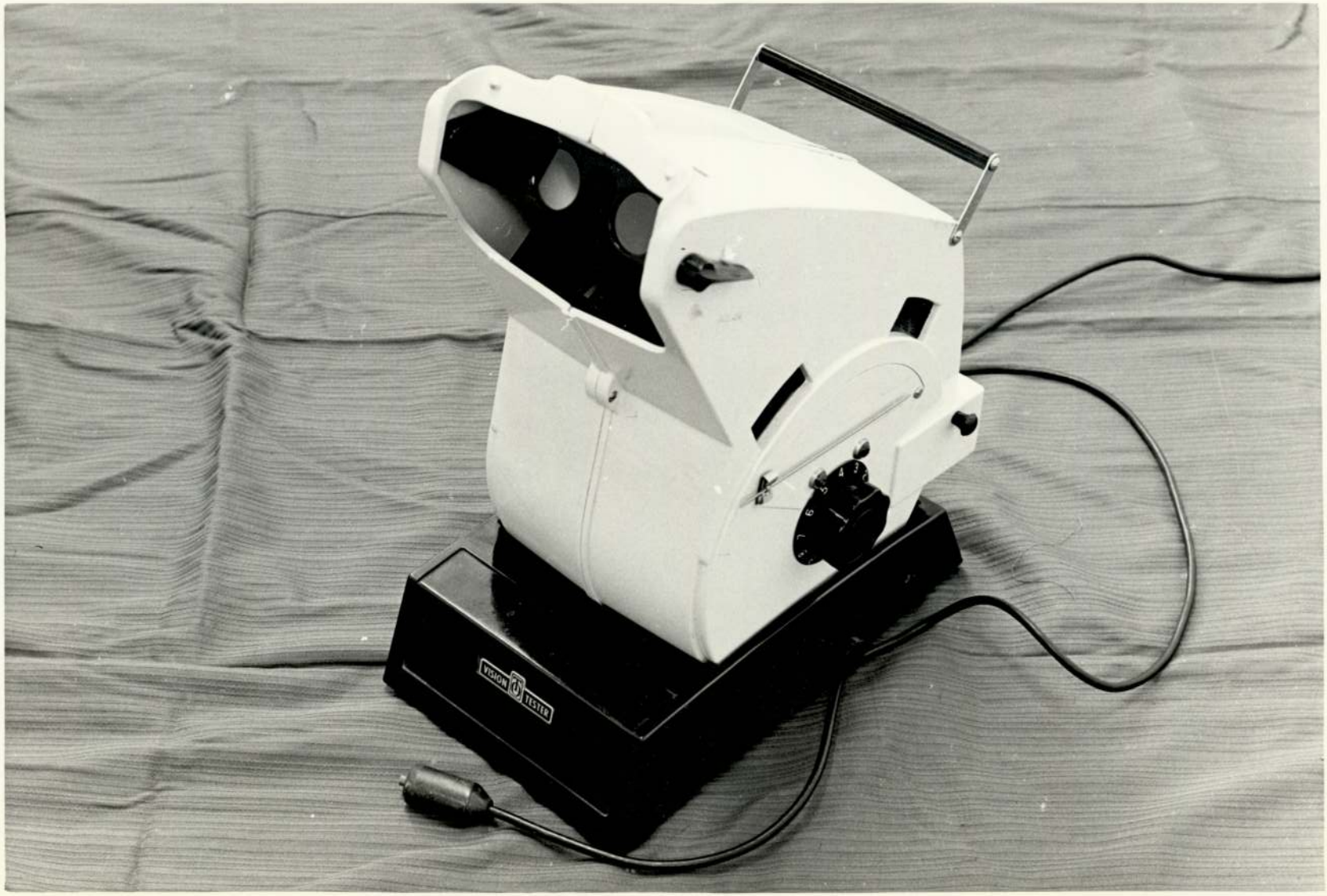
There are fourteen tests:

1. A letter acuity chart for distance monocular visual acuity.




B&O




VISION TESTER

2.& 3. Fogged acuity charts to detect hypermetropia.

4. Distance vertical heterophoria test.

5. Distance lateral heterophoria test.

6. Stereopsis for distance.

7. Simultaneous foveal vision with fusion, for distance.

8 & 9, Near visual acuity for each eye separately, at 33 cm.

10. Near vertical heterophoria.

11. Near lateral heterophoria.

12. To test the ability of the subject to accommodate 5 dioptres.

13. Binocular visual acuity and simultaneous foveal vision.

14. Stereopsis at near visual distance.

The manufacturers provide only one general template to indicate referral.

THE TITMUS OPTICAL COMPANY PROFESSIONAL VISION TESTER

The tests are mounted on a revolving drum, which, in the model used for these experiments, could be revolved electrically. The slides are trans-illuminated.

There are twelve test slides:

1. Distance acuity of both eyes, measured by Landolt ring chart.

2. Distance acuity right eye.

3. Distance acuity left eye.

4. Distance stereopsis.

5. Colour discrimination at 6 metres.

6. Distance vertical muscle balance.

7. Distance lateral muscle balance.

8. Demonstration slide, a stereo-photograph designed to put the subject at ease, reduce tension and stimulate interest.

9. Near visual acuity of both eyes.

10. Near visual acuity right eye.

11. Near visual acuity left eye.

12. Near lateral phoria test.

A set of templates is supplied with the instrument. The standards for referral are those based on the work done at Purdoe University, and are the same as those used with the Ortho-Rater.

APPENDIX II

CODING INFORMATION FOR PUNCHED CARDS

The punched cards used in this investigation were I.B.M. 80 column cards. Each column has the capacity to carry one code symbol, either a letter or a digit. It is preferable to code in digits as this requires only one hole to be punched. Letters require two holes; the alphabet is divided into three parts, the first hole determines from which segment of the alphabet the code letter is taken, and the second hole determines which position the letter holds in that particular segment. As there are two holes, the cards have to be sorted twice when sorting into letters, this is time consuming.

The following list shows the codes and column allocations for transferring screening results to a coding sheet.

Column No.	Information	Code
1	Blank	
2. 3.	Year of test e.g. 1966	66
4 - 7	Enrolment number	
8 - 18	Name of student - Eleven columns are alloted to the name, the first two for initials, the rest for the surname. Where the name is too short to fill the entire space, the remaining columns are left empty. Where the name is too long, the first nine letters are entered, this usually being enough for identification purposes.	
19	Sex: Male	M
	Female	F
20 - 22	Course: Using the same code as the University of Birmingham's registry.	
23 - 24	Year of birth e.g. 1948	48
25	Screening technique used:	
	Manual	1
	"Mavis"	2

Column No.	Information	Code
26	Glasses worn:	
	No glasses	1
	Glasses	2
	Reading glasses	3
	Contact lenses	4
	Glasses not brought	5
	Reading glasses not brought	6
27	Date last test:	
	Under one year ago	1
	Over one year	2
	Never	3
28	Hypermetropia L.E.:	
	T.I.B. through +2.00 D spheres.	
	Normal, letters not seen	1
	6/12 letter read	2
	One 6/7.5 letter read	3
	Both 6/7.2 letters read	4
29	Hypermetropia R.E.:	
	As above	
30.. 31.	Visual acuity R. & L.	
	T.I.B. viewed directly.	
	Coded as for hypermetropia, except	
	that "normal" will be both 6/7.5	
	letters read	
32. 33	T.I.B. with 3^{Δ} base in prism	
	Coded as above	
34	Distance Vertical Phoria:	
	("Mavis" only)	
	Coded as the number recorded on the	
	record card	1 - 8

Column No.	Information	Code
34 (continued)	Distance Vertical Phoria:	
	("Mavis" only)	
	No line seen	9
	No numbers seen	0
35	Distance Lateral Phoria:	
	("Mavis" only)	
	Coded as the letter recorded	
	on the record card	K - Q
	No arrow seen	X
	No letters seen	Y
36	Near Test R.E.	
	Not seen	1
	Rose Dale	2
	Rose	3
	Dale	4
37	Near Test L.E.	
	Not seen	1
	Lead Sore	2
	Lead	3
	Sore	4
38	Convergiometer:	
	Coded as letter read from scale	R-2-O-D
	One arrow seen	I
39	Colour Vision:	
	Four plates are shown, the first	
	is demonstration plate. Of the	
	remaining three:	
	All correct. Normal.	1
	One incorrect, slightly defective	2
	Two incorrect,severely defective	3.

Column No.	Information	Code
40 - 42	History and Symptoms: Recorded on three columns so that several correlated symptoms can all be recorded. It also enables pathological conditions to be specified by position.	
40	A	
	No complaint	1
	Recent Test	2
	Recent R_x	3
	Request Test	4
	Poor Motility	5
	Convergence insufficiency	6
	R.E.	7
	L.E.	8
	Both eyes	9
41	B	
	Blurred vision	1
	Diplopia	2
	Orthoptics	3
	Operation	4
	Trauma	5
	Pathological condition	6
	Bifocals	7
	Amblyopia	8
42	C	
	Headaches	1
	Cornea	2
	Iris	3
	Lens	4

Column No.	Information	Code
42(continued)	C	
	Retina	5
	Adenexa	6
	Strabismus	7
43	Cover Test (Phorias):	
	Divergent under cover:-	
	Quick response	1
	Medium response	2
	Slow response	3
	Convergent under cover:-	
	Quick response	4
	Medium response	5
	Slow response	6
44	Cover Test (Tropias):	
	L. Divergent strabismus	1
	R. Divergent strabismus	2
	Alt. Divergent strabismus	3
	L. Convergent strabismus	4
	R. Convergent strabismus	5
	Alt. convergent strabismus	6
	Vertical strabismus	7
45	Assessment:	
	✓ Passed	1
	S Satisfactory	2
	R Referred to own optician	3
	X Referred next term	4
	X Referred this term	5
	XX Referred immediately	6

The code for letters on the test chart was made easy by the fact that, of the two 6/7.5 sized letters for each eye, one was much more easily read than the other:

i.e. R. E. - L & N

L. is more easily read than N and so if only one of the 6/7.5 line was read, it was invariably the L, hence scaling and coding was made possible.

APPENDIX III

QUESTIONNAIRE

UNIVERSITY HEALTH SERVICE

I am analysing the efficiency of the visual screening technique used at the freshmen's medical examination and would be glad of your co-operation in answering this questionnaire. Please return it to me through the internal University post in the envelope provided.

R. H. Bolton
University Medical Officer

Please put a tick by your answer.

Since the eye test you had on entering the University of Birmingham:-

(a) Have you had your eyes tested:-

No	1	73
Yes, at the University Health Centre	2	
Yes, elsewhere	3	

(b) Have you had glasses or contact lenses prescribed for you:-

No	1	74
Yes	2	

(c) Have you had any of the following:-

Frequent headaches	1	75
Eye strain	2	76
Blurred vision	3	77
Double vision	4	78
Eye injury	5	79
Eye operation	6	80

(d) Comments:-

REFERENCES

AGARWAL, L.P. and DAS.P.C. An evaluation of the Orthorater.
Orient. Arch. Ophthal. 2. 101-110 May 1964.

ALLEN, N. Physical culture in Amherst College.
Stone and Hase 1869.

ANON. Vis. screening in schools. Recommendations of the N.S. for the prevention of blindness.
Sight-saving Rev. 31. 50-58 Spring 1961.

BENNETT, A.G. Ophthalmic test types. A review of previous work and discussions on some controversial questions.
B.J.P.D. 22. 238-271. No.4. 1965.
Ophthal. List.

BETTS, E.A. Prevention and correction of reading difficulties.
Row, Peterson and Company. 1936. Evanston. Ill.

BLACKHURST and RADKE. School vis. screening in the State of Michigan.
Sight-sav. Rev. 34. 8-15. Spring 1964.

BLUM. H.L., PETERS, H.B., BETTMAN, J.W., FELLOWS, V. Jnr., JOHNSON, F.
Design and eval. of a vis. screening programme for Elementary School Children.
Am. J. Pub. Health.

BLUMBERG, M.S. Evaluating health screening procedures.
Operations Research 5. 351-360 June 1957.

BOLTON, R.H. T.I.B. as part of a screening test for Univ. freshmen.
Trans. of the Internat. Op. Congress 1951 pp.303-308.

CAL. STATE DEPT. OF PUB. HEALTH and CAL. STATE DEPT. OF EDUCATION.
A guide for vis. screening of school children in the public schools of California; Recommendations.
Sacramento, Cal. State Dept. of Education 1953. (Peters H.B. 1961).

CHAPANIS, A. Ergonomics. 10.5. 557-577. 1967.

CINOTTI, A.A. and SIEGEL, R. Eye health screening in N. Jersey.
Sight-sav. Rev. 34. 143-151 (Fall. Aut. 1964).

COLLEN et al. Automated multiphasic screening and diagnosis.
Amer. J. Public Health. 54. 741-50 May 1964. Index Medicus '64.

COX, B.J. Vision care and development program of Langley School District.
Canad. J. of Optometry. 28. pt.4. 93-95. 1967.

CRANE, M.M. We study testing children's eye-sight.
Child. 15. 15-17. 1950.

CRANE, M.M., SCOBEE, R.G., FOOTE, F.M. and GREEN E.L.
Study of procedures used for screening Elem. School Children for visual defects. Referrals by screening vs. ophthalmological findings.
J.Sch.Health. 23. 44-49. Feb. '53.
Am. J. Public Health. 42. 1430-1439 Nov. 1952.

CUTLER and DAVEY. Visual screening. Motor Show 1961. Racing Car Show '61/'62. The vision of the 2 groups of drivers.
Brit. J. P.O. 22. 114-119. No.2. 1965.

DISKAN, S.M. Eye problems in children.
Postgrad. Med. 34. 168-78. Aug. '63.

DISKAN, S.M. A new vis.screening test for school children. A prelim. test of 799 school children.
Am. J. Ophthal. 39. 369-374 March '55.

FINNEY, D.J. Screening Processes: Problems and Illustrations.
Contributions to Stats. ed. C.R. RAO p.73-100.

GALLAGHER, J.R. and GALLAGHER, C.D. Colour vision screening of pre-school and 1st grade children.
Arch. Ophthal. (Chicago) 72. 200-211 Aug. '64.
Medlars.

GENTILE, J.F. and JOHNSTONE, P.W. A comparison of 2 school vision screening tests.
Amer. J. Ophthal. 52. 511-515 Oct. 1961.

GORDON, ZEIDNER, ZAGORSKI and UHLANER. V.A. measured by wall charts and orthorater tests.
Am. J. Ophthal. 37. 699-705. May 1954.
J. App. Psy. 38. 54-58. 1954.

GUTMAN, E.B. School vision screening. A comparison of 2 methods.
Sight-sav. Rev. 26. 212-219. Winter 1956.

HATHAWAY, W. Education and health of the partially seeing child.
Columbia Univ. Press, IVth Edition 1959. p.22.

HIRSCH, M.J. The variability of retinoscopic measurements when applied to large groups of children under Visual screening conditions. Am. J. Optom. 33. 410-416. 1956.

HIRSCH. M.J. Skiascopic procedure for visual screening surveys.
Am. J. Optom. 27. 587-591. Dec. 1950

HUNT. W.T. Jr. and BETTS E.A. Visual problems: certain assumptions and data.
Am. J. Ophth. 25: 1084 - 1094. 1942.

IMUS. H.A. Comparison of Orthorater with clinical ophthalmic examination.
Trans. Amer. Acad. Ophth. and Otalaryn. 54. 69-71. Sept-Oct. 1950.

JOBE. F.W. Bausch and Lomb Magazine. Vol. 20. 2. p.6.

KUHN. H.S. Ophthalmic screening procedures.
Trans. Ophthal. Soc. Aust. 17. 106-110. 1957.

KUSHNER. F. Analysis of screening scores.
Optom. Wkly. 44. 967-969. June 4. 1953.

LANCASTER. W.B. et al. Standards for referral of school-children for an eye examination.
Am. J. Ophthal. 37 (5) 710-718. 1954. Peters H.B. '61.

LEVERETT. Vision test performance in schoolchildren.
Amer. J. Ophthal. 44. 508-519. Oct. 1957.

LIPPMANN. O. Eye screening.
Arch. Ophthal. 68. 692-706. Nov. 1962.

MACRAE. D.S. San Jose's approach to occupational vision testing. Publ. Personnel Rev. 17. 129-131. 1956.

MARTIN W.F. Results of a research study on the vision of athletes. Optom. Wkly. 55. 23-27. July 1964.

MICHIGAN DEPT. OF HEALTH. Pre-school hearing, vision screening results revealed.
Mich. Med. 64. 958. Dec. 1965. Index Medicus '66.

MORGAN. A. L., CRAWFORD J.S., PASHBY and GABY.
A survey of methods used to reveal eye defects in schoolchildren. (The Toronto Study). Trans. Canad. Ophthal. Soc. 14. 125-139.1951.

MURPHY and THYNG. School vision screening.
J. Indiana Med. Ass. 50. 1345-1347. Oct. 1957.

NEYMAN. J. Outline of statistical treatment of the problem of diagnosis. Pub. H. Report. 62. 1449-1456. Oct 3 1947.
Collen et al. Am. J. Pub. H. 1964.

OAK. L. and SLOANE. A.E. The Betts Visual Sensation and Perception Tests. Arch. Ophth. 22. 832-843. 1939.

PETERS G.A. An appraisal of visual defects of children in Indiana. J. Indiana. M.A. 31 237-238 1938.

PETERS H.B. Vision screening with a Snellen Chart.
Am. J. Optom. 38. 487-505. Sept. 1961.

POWELL and MIRSKY. Lions Clubs join optometrists in a vision screening program. Optom. Wkly. 55 22-6. 16 Jan 1964.
Medlars.

REESE. G.A. School vision screening. An evaluation of the Mod. Clin. Tech.
Trans. Pacif. Cst. Oto-Ophthal. Soc. 45. 241-271. 1964.

REPORT OF COUNCIL ON PHYSICAL THERAPY. Betts Charts not acceptable.
Report of Council on Physical Therapy. J.A.M.A. 113. 937-938. 1939.

ROBERTS J. Comparison of two vision screening devices.
Pub. Health Service Publn. 1000. Series 2 No. 1.
Washington U.S. Dept. of Health, Educ. & Welfare. Pp33 1963.

ROBINSON. H.M. Analysis of four visual screening tests at grades 4 - 7. Am. J. Optom. 30. 177-187. April 1953.

ROSENBLOOM. A. Junr. A critical Evaluation of visual diagnostic materials.
Elem. School. J. 56. 27-31. 1955.

SCHWARZ K. Screening tests.
Brit. J. Clin. Pract. 19. 421. Jul. '65. Index Medicus '65.

SLOAN L.L., ROWLAND W.M. and ALTMAN. Comparison of three types of test target for the measurement of V.A.
Quart. Rev. Ophth. 8.1. 4-16. 1952.

SLOANE. A.E. Massachusett's vision test.
Arch. Ophth. 24. 924. Nov. 1940.

SLOANE and GALLAGER. Vision of adolescent boys.
Am. J. Ophthal. 33. 1746 - 1750. Nov. 1950.

SLOANE A.E. and GALLAGER J.R. A comparison of Vis. screening tests with clinical examination results.
Am. J. Ophthal. 35. 819-830. June 1952

SLOANE and SAVITZ. Vision screening.
Int. Ophthal. Clin. 3. 803-9 Dec. '63. Medlars.

SNYDER. C. The professor and the Jakes cleanser
Arch. Ophthal. 72. 441-443. Sept. 1964. Ophthal. Lit.

SPACHE. G. Testing vision.
Education. LIX 623-626. June, 1939.

SPROUL. A. Influence of conditional probability on the effectiveness of mass-screening programs.
J.A.M.A. 196. 315-318. 25th April 1966. Index Medicus '66.

THORNER and REMEIN. Principles and procedures in the evaluation of screening for disease.
Pub. Health Monog. No. 67. Washington D.C. Govt. Ptg. Office 1961. Collen et al. Am. J. Pub. Health '64.

TURVILLE A.E. Some recent expts. in infinity balance.
Trans. of the Internat. Optical Congress 1951. P.299-302.

UNGAR. P.E. Drivers visual screening.
Oph. Optic. 2. 513. 1962a.

UNGAR. P.E. Vision screening at Brands Hatch.
Oph. Optic. 2. 827. 1962 b.

VAN WOERKOM AND BRODMAN. Statistics for a diagnostic model.
Biometrics. 17. 299-318. June 1961.
Collen et al. Am. J. Pub. H. 1964.

VECCHIO. T.J. Predictive value of a single diagnostic test in unselected populations.
New England J. Med. 274. 1171-3. 26. May '66.
Index Medicus '66.

WEINHOLD. H. Rapid vision tester for mass investigations.
Yena Rev. 9 pt. 1. 87-89. 1964.

YASUNA. E.R.and GREEN L.S. An evaluation of the M.V.T. for visual screening of school children.
Am. J. Ophthal. 35. 235-240. Feb. 1952.