How to use the Ishihara test

This is a screening test only. Before using it you should be familiar with its limitations since incorrect use and interpretation are a common cause of misdiagnosis.

Some theory

The American National Institute of Standards and Technology (NIST)[1] was formerly known as the National Bureau of Standards (NBS)[1]. They developed a standardized method of measuring colour difference called NBS units relevant to commercial colour production. Colorimetric data for the Ishihara plates has been published using NBS units.[2] The physics of calculating colour spaces and colour differences is complex. For more information visit the CIE website[3] or the NIST.

One NBS unit of colour difference is equivalent to around 5 'just noticeable differences' (JND) for a standard observer under ideal viewing conditions. The exact number varies with the direction in colour space. In colour normal observers the 'just noticeable difference' increases with age. For most Ishihara plates the figure background difference is 25 to 40 NBS units, although it varies from as little as 10 in one plate up to 55 units in another.[2] For plates designed to be deliberately ambiguous, known as transformation plates, the difference between the alternate foregrounds is also relevant, especially in that part of the figure that causes the ambiguity. The plate that should read 74 can be misread in a variety of ways, often as 21 or sometimes 91, since the angular components of the target that allow it to be correctly identified do have as little as 10 NBS units contrast and this can be greater than the physiological 'just noticeable difference' in some older observers. This means errors in some plates are expected in colour normal observers with overall poor discrimination (cataract, optic neuritis, toxic maculopathies where fixation is preserved, etc.) and it also means that whenever colour contrast is reduced by poor test conditions (old faded plates or incorrect illumination) the false positives occur in some plates more than others.

History of the test

Shinobu Ishihara (1879 to 1963) graduated from medicine in 1905 on a military scholarship and immediately joined the army as a doctor, serving mainly as a surgeon. He later changed specialties to Ophthalmology. In 1908 he returned to Tokyo University where he dedicated himself to ophthalmic research. In 1910 he became an instructor at the Army Medical College. There, in addition to seeing patients, he conducted research on "battlefield ophthalmology" and how to select superior soldiers. In 1912 he went to Germany to further his studies in Ophthalmology and in 1915, after the outbreak of war, he returned to Tokyo. (Japan entered WW1 as an ally of Britain on condition it could take Germany's Pacific territories). There he worked as an instructor in the Military Medical School where he was asked to devise a test to screen military recruits for abnormalities of colour vision. His assistant was a physician with X linked

anomalous colour perception who helped him test the plates. The first charts were hand painted by Ishihara in watercolours using hiragana symbols – the most "Japanese" of the three Japanese scripts. These symbols were derived from Kanji around the 9th century. Previous versions of European and Japanese pseudo isochromatic plates existed but Ishihara's plates gave more reliable results. In 1917 he made a set using Arabic numerals that he called the "International Edition" but few copies were sold. In 1922 Ishihara became a Professor at Tokyo University. In 1929 at the 13th International Congress of Ophthalmology in Holland the International Edition was recommended for testing naval personnel and air force pilots. In 1958 the "Law of School Health" in Japan required that a 'colour blindness' check be done as part of an overall health check on young school children. The Ishihara test was designated as the official test for this purpose, and after this the test achieved widespread use within Japan. Since then Ishihara's charts have been used throughout the world and are now the commonest screening test for colour vision anomalies.

How to ensure reliable results

For correct test conditions ensure (1) correct illuminance (350 to 600 Lux) and correct wavelength (either natural daylight, which is a south facing window in the southern hemisphere, or more commonly artificial daylight, 4 daylight fluorescent globes within 1.5 meters above the page). We use a commercially available artificial daylight desk lamp. (2) correct viewing geometry – line of sight at 45 degrees to axis of illumination, distance at 35 to 45 cm (3) correct test - printed pigments fade at different rates with age, so tests older than 10 years seldom look the same as a new test (4) correct method – avoid glare, provide optical correction, avoid highly coloured surrounds, allow up to 5 seconds per page view. When screening for occupational standards consider covering the page numbers at the bottom with a sticker and presenting the pages out of order to deceive cheats who memorize the correct responses.

The different types of plates in the test

There are two editions - a 24 plate series and a 38 plate series. It is best to use the larger series because there are relatively few reliable plates in the smaller series. Both sets consist of two groups of plates - a group for those who are numerate that starts at the front of the book, and a group for innumerates that is colorimetrically identical and intended to proceed in reverse order starting at the back of the book. The group of plates for innumerates are seldom used because they are almost impossible to score consistently – and the same would be true for a meandering black line on a plain white page – the extent to which one wanders off the line can be affected by any number of irrelevant variables. From a colorimetric perspective both the numerate and innumerate sets consist of four different types of test plate, and this is true for both the 38 and 24 plate series. The comments that follow related to the numerate plates since the innumerate ones should never be used for a test where the results are important. In both series plate 1 is for demonstration. If the subject fails this plate the visual function is too poor to proceed. The comments that follow refer to specific plates

in the 38 plate series (with corresponding numbers for the 24 plate series in brackets).

The four test designs are:

1.**Transformation plates** - anomalous colour observers give different responses to colour normal observers. These are the plates numbered 2 to 9 inclusive, (or 2 to 7)

2.**Disappearing digit,** or vanishing plates - only the normal observer is meant to recognize the coloured pattern. These are plates 10 to 17 inclusive in the 38 plate series, (or 8 to 13).

3.**Hidden digit** plates - only the anomalous observer should see the pattern. These are plates 18 to 21 inclusive, (or 14 and 15).

4.**Qualitative** plates - intended to classify Protan from Deutan and mild from severe anomalous colour perception. These are numbered 22 to 25.

Plates 2 to 21 are made from a selection of only 9 inks, whilst plates 22 to 25 are made from just 6 inks. It should be noted that the colorimetric task is slightly different in plates 14 to 17, where the foreground green inks are less blue than in the earlier plates while the background has a lot more orange and yellow than in the earlier plates.

Which plates to use – the 'ideal subset.'

Of these 4 types **the only reliable designs are the transformation and disappearing digit types**, (2 to 17 inclusive in the 38 plate series, or 2 to 13 in the 24 plate series). I will refer to these as the *ideal subset*. The score is therefore out of either 16 (38 plate series) or 12 (24 plate series).

The average false positive rate is 8%[2] in a general clinic population. Within the *ideal subset* the false positive rate varies from plate to plate and increases with age from late teenage years. All the plates with 2 numerals are more difficult than those with 1 numeral given the same colorimetric task, and plate No.9 is notoriously difficult.

The false negative rate varies with the nature of the anomalous perception but is low (our experience suggests less than 5% per plate amongst anomalous observers).

Cut-Off Scores

Several sophisticated analytical approaches have been used to determine the best cut-off score to apply in population screening tests. Using Bayesian probability and a plot of sensitivity Vs. 1-specificity, known as a receiver operator characteristic (ROC curve), a cut-off score of 4 or more errors is often recommended in a young male population.[4] The confidence of this end point improves if the *ideal subset* is used. If the full numerate set of plates from 2 to 25 are used in a general population then a cut off at 2 or more errors could fail up to 45 of 100 colour normal older males of whom 37 would be false positives. If the

'ideal subset' of 16 plates from the 38 plate series is used and a cut off score of 4 errors or more errors is the fail criterion then we should expect to fail 12 of every 100 randomly selected males, of whom 8 are true positives and 4 are false positive results. This relatively small group can easily be submitted to further testing – such as a D15 or anomaloscopy test – to divide the true positives from the false. Since the rate of false negatives is not significant unless the cut off score exceeds 6, it may be quite satisfactory to set the cut off score at 5 or even 6 if the group being tested have an abnormally high prior probability, such as older patients. This decision is always a balance between the risk of false positives and false negatives and the purpose for which the test is being used. If it is essential not to miss any anomalous observers a cut-off at 3 or more may be preferred, but that will generate a larger number of false positives who have to be filtered out with further testing.

When a non ideal subset of plates is employed the pass/fail criterion will be different. The disadvantages of the 24 plate series are that there are only 12 reliable plates in the entire test, and that the cut off score of 3 or more is not as valid an end point as using a score of 4 or more with a 16 plate ideal subset.

If the Hidden Digit plates are included the cut-off score must be increased to allow for the predicted false positive results, but there is no reliable way of determining what the cutoff should be since the rate of false positives is very age dependent. The highest false positive rate for these plates occurs in young adult males – the very group we are most likely to use the test in.

The distribution of normal and anomalous observers varies enormously between the female and male populations. In Bayesian terms the prior probability of a true positive result is lower in females, so the cut-off needs to be lower, since the relative importance of a false negative result is greater when there are so few true positive cases amongst the test subjects. Unfortunately the carrier condition may confuse matters and the affected female retina is possibly a mosaic of different anomalous pigments, so the clinical characterization is likely to be less reliable. False positives tend to dominate the population who fail the test so a higher cut off may be preferred but an alternative is to avoid the Ishihara test altogether in females because no one is sure how to interpret the result in the absence of other tests.

Hidden Digit And Qualitative Plates

The four Hidden Digit plates vary enormously with age group and degree of anomaly, but unlike all the other plates are misread more often by those in the 20 to 40 age group and less in older observers. Anomalous observers are intended to read 5, 2, 45, and 73 respectively. Normals are intended to read no number but in one study of healthy young psychology students in their 20's 40% read 45 on the third plate, and false positive rates averaging 20% apply to each of the other plates in this series within the15 to 30 year age group.[3] This is precisely the age group who are most likely to be getting screened with an Ishihara test for employment related purposes.

Qualitative plates consist of 6 inks with a pink figure intended to be confused with grey by protans and a purple-red figure intended to be confused with the grey background by deutans. In one study 1 out of 18 outright protanopes responded as intended and 6 out of 18 outright deutanopes.[3] Since both types of plate are so unreliable they should not be used.

Two common errors in interpreting the results.

There are three common errors of interpretation. The first is to assume that an error detected in just the red / green axis is sufficient to determine overall colour confusion in every axis. This is particularly relevant to those acquired conditions like optic neuropathy that primarily produce confusion in a blue /yellow axis. The second is to overestimate the dynamic range of the test. Normal observers can distinguish a total of up to 10 million colour stimuli[5] using additive mixes but for subtractive colour mixes of the type presented in offset printing process the dynamic range is much less, and varies with gloss, paper substrate, and other material choices. The maximum dynamic range that could be produced along the red / green axis using high quality matte printing analogous to the Ishihara plates is several times the 55 units used in the easiest Ishihara plate, whilst the maximum difficulty could be as little as ¼ of one NBS unit if one 'just noticeable difference' for a normal observer was used. The third is to misunderstand the information content in different types of measurements - nominal, ordinal, interval or ratio. The Ishihara test provides nominal and ordinal data only. It is an error of logic to interpret the result of an ordinal test as if it was an interval score. The house number in a street of houses built on different sized blocks of land does not enable you to calculate the distance of each house from the end of the street. If the Ishihara test provides useful interval data in acquired dyschromatopsias then it must do so by some mechanism other than the interval distribution of colorimetric tasks. One possible hypothesis is that is could be detecting differences in signal latency between the parvocellular and magnocellular retino-cortical pathways. For colorimetric measurements of patients with optic neuritis an interval blue / yellow test, or the 100 hue test would be more appropriate.

Conclusion.

The Ishihara test is a screening device to quickly sift through a large population of males and sort them into those who are normal colour observers and those who MAY have a colour vision anomaly. It is an excellent test for this purpose but it does select false positives and cannot quantify the colour anomaly accurately or grade the severity. Other tests must be used for these tasks.

1. NIST. http://www.nist.gov/index.html
2. Lakowski, R. Colorimetric and Photometric Data for the 10th Edition of the Ishihara Plates. British Journal of Physiological Optics 1965;22:195-207.
3. CIE. http://www.cie.co.at/

4. Aspinall P.A. & Hill A. R. Decision analysis for colour vision testing, In Verriest G, editor. Colour Vision Deficiencies VI, Volume 6 (Proceedings of the IRGCVD Symposium, Berlin, 1981). Heidelberg: Springer; 1982.
5. Wyszecki, G. Color. Chicago: World Book Inc; 2006.