Online games for colour deficiency data collection

Luca Armellin[1], Alice Plutino[1], Alessandro Rizzi[1]

[1]MIPS Lab, Department of Computer Science, Università degli Studi di Milano, Milano

Contatto: Luca Armellin, luca.armellin@studenti.unimi.it

Abstract

Trying to investigate the role of edges in the color perception of color vision deficient people requires complex unstandardized setups, possibly leading to longer and more challenging tests. Besides the fact that only less than a tenth of the western population shows some form of color blindness, adding inherent complexity to test setups may likely worsen the scenario and further reduce the availability of test subjects. A possible solution might come from the development of web-based tests, which on the one hand makes it easier for a subject to be enrolled in a study, on the other introduces variability in the form of different devices, environmental and viewing conditions. Not being able to directly monitor a subject also makes it impossible to evaluate its attention and motivation, which likely have a role in the accuracy of the responses given.

In this study, we are investigating the usage of a specifically designed web-based game as a source for larger amounts of data related to color perception; the idea is to exploit the potential of games to become viral and collect large amounts of data with little effort while at the same time addressing variability by means of averaging the outcomes over a large set of test subjects. Developing simple and engaging games might also solve the issue of low motivation and attention, giving to the test subject a reward in the form of entertainment and progress in the game.

Keywords: color deficiency, color blindness, color perception.

Introduction

Color vision deficiency is described as an impairment of the ability to distinguish some colors according to the type of vision defect. This issue has a higher incidence in males (about 8.8%) than in females (about 0. 4%) (Birch, 2012; Hunt and Carvalho, 2016). Screenings and diagnosis of the type and severity of a color vision deficiency can be of great help for the individual, especially in scholarly or young age, when very often the inability to distinguish color could cause discomfort and frustration. Furthermore, a reliable assessment of color vision deficiency can help people in planning their future career or finding creative solutions to overcome their limits.

Color vision deficiency is a genetic or more rarely acquired condition caused by the absence or alteration of one or more cone types in the retina. In general, the total lack of ability to distinguish colors is called monochromacy or color blindness, while the term color deficiency is referred to dichromacy, caused by the absence of one type of cone, and to anomalous trichromacy, caused by the alteration of one cone type. In general, the type of color deficiency depends on the type of cone which is absent or altered: protanopia and protanomaly refer to cones L (red), deuteranopia and deuteranomaly refers to cones M (green) and tritanopia and tritanomaly refer to cones S (blue).

The most common type of test are the Pseudo Isochromatic Plates (PIP), like the Ishihara test. In these tests the observer should identify a colored symbol embedded in a colored background and colors are chosen to be easily distinguished by trichromats since they lay on color deficiency confusion lines (Judd, 1945). The great success of paper and digital versions of these tests relies on their easy applicability, since using a few tables it is possible to assess the presence of color deficiency and have a preliminary screening of the type of defect. Tests like Ishihara plates are widely used in clinics and occupational environments and present a high sensitivity in screening congenital red and green deficiencies, while this test is not suitable to screen tritanopia/tritanomaly (Cole, 2007) (Rodriguez-Carmona *et al.*, 2021).

Another important and diffused family of color vision screening tests are the arrangement tests. In arrangement tests, the observer is required to arrange color samples by similarity in a sequential color series. Tests from this family, like the FM 100-hue test of the Farnsworth Panel D-15, are very intuitive and easy to administer, but require the observer's patience and concentration, hence they are less suitable for young children. Those tests can be useful to diagnose the type of dichromacy but perform less well to diagnose the anomalous trichromats (Rodriguez-Carmona *et al.*, 2021).

Among other less diffused families of color deficiency diagnosis tests like lantern tests, CAD test or computer-based test, it is mandatory to name the anomaloscope. The anomaloscope is the standard instrument to diagnose color vision defects and classify the defect type. It is an optical instrument, where the observer manipulates stimulus control knobs to match two colored fields in color and brightness. Thanks to this test, it is possible to characterize not only color deficiencies but also anomalous trichromats. Despite its high precision and reliability, the anomaloscope is very difficult to use since it requires an extensive training of the examiners, constant supervision and calibration (National Research Council, 1981). Considering all the specific characteristics of the most used color vision assessment tests, today it could be very hard to obtain a consistent set of data to develop the research on color deficiency or just to raise the awareness of common people to color vision issues. Specific tests can require concentration, cause stress and frustration on the patients because of their clinical nature and be unsuitable for young children or people with behavioral problems.

The development of web-based tests could be a preliminary solution to gamify classical clinical tests and collect large amounts of data to make research and improve the current tests and diagnosis methodologies. Certainly, this solution will never substitute a clinical diagnosis, but it could become an alternative to test in large-scale color vision issues and to make screening in gaming conditions.

Following this idea in this paper we define some tips and tricks to develop an online color vision test based on our experience and we present a first practical example which helped us to test a large number of observers on the role of edges in color vision assessment.

Tips and tricks to develop online tests for color deficiency data collection

The development of online games to test color vision defects has several pros and cons which must be considered before starting to design the game. An online game is surely suitable to reach and enroll a large number of subjects in the study and collect a large amount of data. Anyway, this advantage also presents several challenges because in developing the online game it is important to define and set up a robust data collection system in a manageable format. Furthermore, reaching several different people with an online game entails device variability and different environmental and viewing conditions. This situation can be listed both in pros and cons of this application since the test results could demonstrate the consistency of a theory in different conditions of observation, especially when the same observer played several times and on different devices. On the other hand, this could lead to errors and approximation linked to the different devices' gamuts of used color spaces. In this context, it is clear that in designing the game some precaution can increase the data consistency and reliability and a further solution could be addressing variability by means of averaging the outcomes over a large set of test subjects. The same observations on the device's variability can be applied also to the observer's attention and concentration. Clearly using online games, it is impossible to constantly monitor the subjects' level of involvement, but the test gamification could also create a positive sense of competitiveness and desire to improve, which will surely increase the number of games played and the concentration without stress of the participants.

Another great advantage of designing online games, differently from apps or other traditional computer tests, is that online tests can be customized and modified at any time and the changes appear immediately in the game. This means that it is always possible to fix errors or introduce improvement during the testing or define different test versions to present to the players at different times.

Considering the observation on the general pros and cons of developing online games to assess and test color deficiency it is clear that it remains difficult to diagnose and classify the type of color vision defect, as well as the severity. Anyway, if the color vision test is designed to test some specific color vision features (e.g., the role of edges in color vision, the importance of shape over the color) designing color vision games could help in collecting data and have preliminary answers and research direction. Furthermore, the entertaining value of the test could be useful to test children or young people and could be of help in involving color deficient people without “torturing” them with long frustrating tests.

Issue	**Tips and trick**
Devices’ variability	Use standard sRGB color spaces Decrease color saturation Outcomes averaged over a large set of subjects
Large amount of data to manage	Set real time data classification system Focus on the most important data Set up an online data visualization environment Define your privacy policy
Test gamification	Design some levels for fun and some levels for test Create players rankings Set up prizes or advantages for the best player

Tab. 1 - Summary of main issues in designing online tests for color deficiency data collection and possible solutions.

In Tab. 1 are reported some possible issues in designing online games to test color vision deficiency and some tips and tricks derived from our experience.

Variability among devices is probably the most difficult issue to address, especially since the simple browser-based games here discussed are intended to be played mainly on smartphones. On one hand it is true that the industry of mobile devices is becoming more attentive to color management, for example Android OS starting from version 8.0 released in 2017 enabled support for wide gamut displays covering DCI-P3 color space along with a simple calibration procedure, Apple is also using wide gamut displays covering almost up to DCI-P3 since the iPhone 7 release in 2016. On the other hand, profiling and calibration procedures are still too simple, like the ones used in Android devices which offer few parameters and are extremely subjective; or completely lacking, like in iOS devices.

Another factor that must be considered is the browser's color management. While Safari for iOS implements gamut mapping using the actual color profile of the device as target color space, in Chrome for Android gamut mapping is implemented using exclusively sRGB as target (or DCI-P3 in devices supporting it), regardless of the actual display profile (EIZO, 2022).

Having to deal with this situation it is mandatory to carefully design the game and the choice of colors, following are some guidelines we came up with:

- Use sRGB color profile since most devices don't support other profiles.
- sRGB is default thus does not need to be explicitly defined while using CSS directives or HTML5 Canvas elements but should be embedded in all images used in the game (W3C, 2022).
- Avoid the use of highly saturated colors since it's more likely for a color near the edge of the sRGB gamut to be rendered unalike between two devices with different non-profiled non-calibrated displays.
- Include in the game at least some levels that are identical or very similar among all the players, so that gathered data can be averaged to address variability and outliers identified and excluded from data analysis.

Large amounts of data gathered from different people and devices are mandatory to offer some reliability and confidence during analysis, but at the same time pose some challenges that should not be overlooked.

The game we developed, which we will discuss further below, is designed to exchange very small sub-kilobyte messages with the hosting server once every few seconds, despite this we experienced a bandwidth usage of over 500mbps just hours after the release, hence care must be given to choosing the right metrics and data to collect, avoiding redundant or superfluous information.

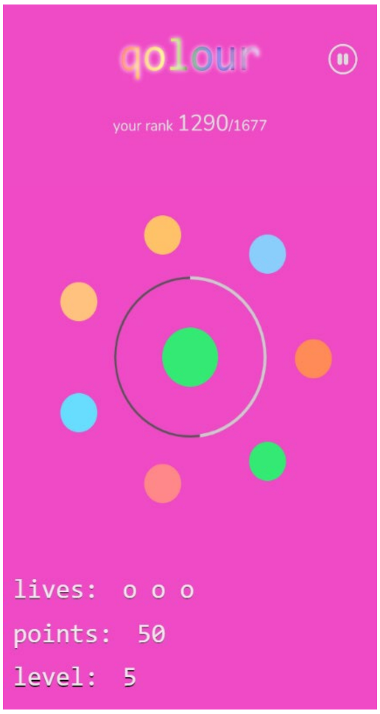
Real time online data visualization and analysis comes in handy when dealing with a dataset in the region of the hundreds of thousands of entries, so that analysis can be performed without downloading and separately processing gathered data.

Given the fact that a user will hopefully play several times, a good practice might be to keep track of its device using cookies and the browser's local storage, thus allowing to keep a record of the player's behavior throughout time, which helps in identifying outliers as sudden changes in playing performance that might be due to environmental factors or learning. But using cookies and local storage requires the end user to explicitly accept a privacy policy, which might keep some people from effectively interacting with the game.

Regarding gamification, a balance needs to be found between the need for meaningful data collection and the player's leisure and satisfaction. Having a game that's too difficult might influence collected information, since errors might be due to the game's difficulty rather than actual phenomenon arising solely from visual perception limits (which we want to test), on the other hand having a game that's too simple might result in a poorly engaging experience. What we did was not to try and gamify a test, but rather design a simple game that could also be used to harvest information; hence we carefully designed some levels which are effectively used for testing and research and others whose main purpose is challenging the player and keeping them engaged.

A practical example

In order to better study the usage of games for data collection we are currently developing a simple game called Qolour. In Fig. 1-a is shown the main interface, as can be seen the player is presented with seven differently colored shapes arranged in a circular fashion, in the center there's a slightly bigger shape surrounded by an animated timer; at the top a global ranking informs the user of its actual position in the global ranking among the other players while at the bottom are displayed its current level, gained points and lives left. The purpose of the game is to press, before the timer expires, on the outer shape with the same color as the central one.



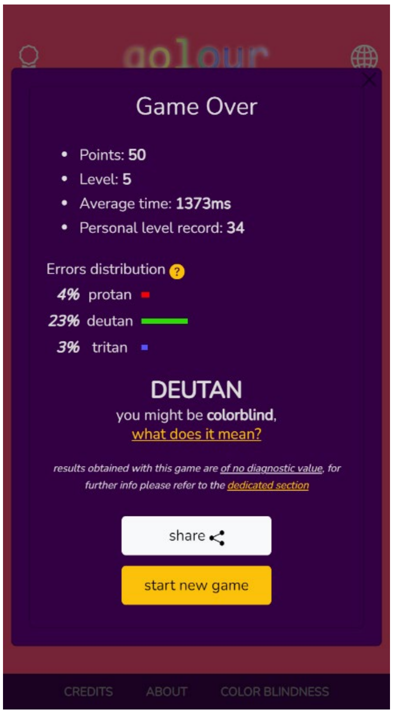
your rank 1290/1677

lives: o o o

points: 50

level: 5

Fig. 1-a – The main interface of the game Qolour as seen on qolour.it.



Game Over

- Points: **50**
- Level: **5**
- Average time: **1373ms**
- Personal level record: **34**

Errors distribution ?

4% protan
23% deutan
3% tritan

DEUTAN

you might be **colorblind**,
what does it mean?

results obtained with this game are of no diagnostic value, for further info please refer to the dedicated section

Fig. 1-b – The modal window showing the overall score and the assessment for the game Qolour.

The timer duration ranges from a minimum of 4 up to a maximum of 8 seconds and serves two main purposes: it increases the difficulty making the experience more challenging and renders more difficult for the end user to use strategies and tools to tamper with the game (e.g., using the browser's debugger to find the exact hex values of the shapes' color). Having noticed that some users had difficulty with tight timings, we decided to implement an adaptive timer, which increases in steps of half a second every time the player makes a mistake, up to 200% the initial time.

The central color, which we will call the *target color*, is generated randomly inside the HSL space, and subsequently converted to RGB, with the Hue being completely random, Saturation bounded between 0.6 and 0.8 and Lightness fixed to 0.5; since this setup is intended to gather data regarding the influence of spatial arrangement on color perception of Color Deficient Observers (CDOs) we decided to keep Lightness fixed in order to remove any variability that might be due to the variation of intensity between different rounds. One of the *outer colors* is the same as the target color, while the remaining six are chosen as to lie on confusion lines corresponding to the three types of dichromats, with the distance (measured in linear-RGB space) from the target color inversely decreasing with levels, so that as the user advances in the game the overall difficulty is increased.

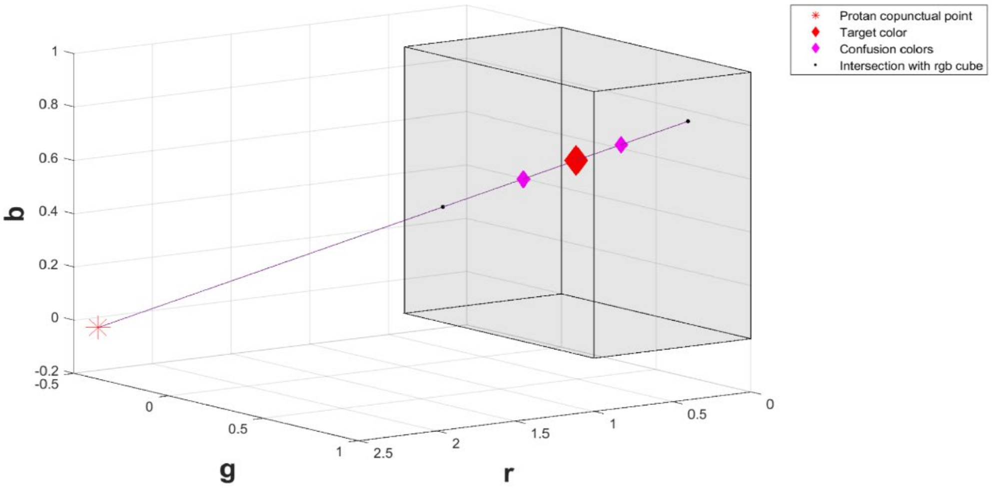


Fig. 2 - Representation of how the confusion colors are computed in linear-RGB space for a given type of dichromacy. The example in figure shows the line originating from the protan copunctual point with the same brightness as the target color, computed in linear-RGB coordinates. Along the segment connecting the two intersections between the ray and the cube, confusion colors can be sampled and gamma-corrected afterwards.

The linear-RGB space is not a perceptually uniform color space, but for practical reasons a choice has been made to compute distance in this space rather than in the perceptually uniform CIELAB space. Confusion lines are lines in the XYZ and linear-RGB spaces (being the transformation between the two linear), so, as shown in Fig. 2, it is easy to compute the coordinates of the two intersections with the linear-RGB cube's faces and a line originating from the copunctual point and intercepting the target color, thus allowing to compute any confusion color simply sampling along the segment connecting the points of the two intersections and later applying a proper gamma correction to transform the color into the sRGB space. In the example depicted in Fig. 3, it can be seen that straight confusion lines turn into nonparametric curves when working in the CIELAB space (but the same applies to other perceptually uniform spaces), thus rendering it difficult to reliably compute confusion colors without sampling the whole space. Since in our setup the target color is, as said, randomly chosen and the outer colors are computed client-side in real time, we faced the need for a fast and reliable computation, at the cost of perceptual uniformity, which is a problem that can easily be addressed later when performing data analysis.

In Qolour, the background color changes with each level and it can randomly be achromatic (having the same Lightness of the target color) or colored (a pseudorandom color computed such that it has the opposite Hue of the target color plus or minus 10 degrees but the same Lightness and Saturation). Data obtained with a gray background are used to roughly estimate a deficiency coefficient that helps in segmenting the users between supposed CDOs and supposed Normal Color Observers (CNOs),

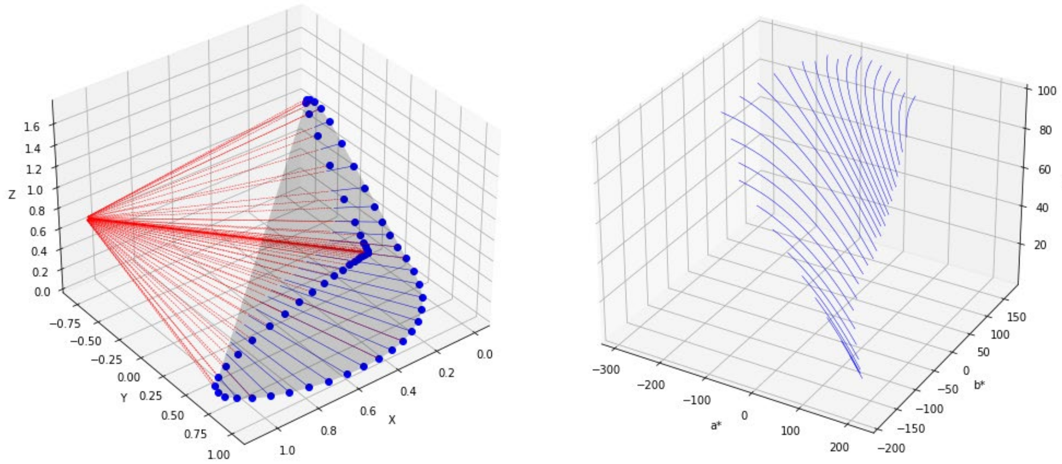


Fig. 3 - Left: protan confusion lines in XYZ space originating from the copunctual point towards the spectral locus (blue dots); the blue portion of the lines is the one falling inside the visible region of the XYZ space. Right: corresponding confusion curves plotted in CIELAB. In this example the copunctual point in XYZ is computed from the xy coordinates as to have Y=y and X+Y+Z=1.

while the differences between the outcomes with colored and gray backgrounds are used to highlight possible contributions of the spatial context in the perception of colors.

Along with changing backgrounds and colors, the shapes can be displayed normally, with a blurred border (box blur via a kernel of approximately 10% the width of the shape) or a solid white border (having a width of approximately 5% the width of the shape). The shapes are always displayed without a blurred or a white border for the first 10 levels, as the role of the first 10 levels is to show the user how to play but are displayed randomly in one of the three configurations with a probability of 1/3 for the next levels. Every time the player picks the right color it gains 10 points and advances a level, every time it fails it loses points proportionally to the current level, if the timer expires it loses a life. The game ends whenever the user reaches 0 points, or 0 lives left; the global ranking is based on the maximum level the user has managed to reach. At the end of the game, as can be seen in Fig. 1-b, a modal window is displayed showing some statistics related to the progress, two buttons to play again or share the current score and a simple error distribution along the three possible directions towards the copuncutal points. As a bonus, a hint suggesting whether the user might be colorblind is shown, with a detailed explanation regarding color blindness and how data can be interpreted. This little assessment serves no diagnostic purposes and is intended solely to better engage the users' attention and give them some information regarding color blindness.

The estimated deficiency is determined based on two scores computed for both deutan and protan directions. The score *Score(d,u)* for each player u (*user*) and deficiency type d (*protan*, *deutan* or *tritan*) Eq. (1) takes into consideration the error rate *E%(d,u)* Eq. (4), which is simply the percentage of errors committed along the confusion lines of deficiency d for the player u, and the median of all the $\Delta E_R(r,u)$ for the player u in all the rounds $R_{d,u}$ for which the picked color lied on the confusion lines of deficiency d Eq. (2). $\Delta E_R(r,u)$ Eq. (3) is computed as the ΔE between the target color $t_{r,u}$ of round r and the color $p_{r,u}$ picked by the player u in the round r, relative to the maximum ΔE between the target $t_{r,u}$ and each of the outer colors c shown to the player u in the current round r. Since most of the errors are made with relatively small ΔE, a logarithm is used in Eq. (1) to emphasize low values and its argument is chosen to obtain a positive value bounded between 0 and 1. The score ranges from 0 to 100, the higher the value the more frequent and/or severe the errors committed along a certain set of confusion lines.

Eq. (1) $$Score(d,u) = E\%(d,u)\cdot\ log_{10}[1+9\ \cdot\ \Delta\widetilde{E_R}(d,u)]$$

Eq. (2) $$\Delta\widetilde{E_R}(d,u) = median[\Delta E_R(r,u)]\ \mid\ r\ \in\ R_{d,u}$$

Eq. (3) $$\Delta E_R(r,u) =\ \Delta E(p_{r,u},t_{r,u})/\max\left(\Delta E\left(c,t_{r,u}\right)\right)\ \mid c\ \in OuterColors_{r,u}$$

Eq. (4) $$E\%(d,u) = 100\ \cdot Errors(d,u)/Rounds$$

A set of 16 control subjects, composed of 8 CDOs and 8 CNOs, have been given a special link to play the game so that we could compare the actual impairment of each one with the estimate resulting from the game. In distinguishing between CDOs and CNOs the metric used in our game resulted in a 100% success rate among this control group.
In a subsequent release a small optional survey has been added to the game interface, asking the users whether they know or think they might have any form of color deficiency. 12 out of 12 players who responded saying they have a Color Vision Deficiency (CVD) were flagged as CDOs by our game, while 32 out of 35 users who responded saying they never experienced CVDs were flagged as CNOs.

Nickname	**Deficiency**	**E%(protan)**	**E%(deutan)**	**Score(protan)**	**Score(deutan)**
ehi.pluto	Protanopia	**23%**	11%	**15.91**	9.86
minicartar	Protanomaly	**10%**	3%	**4.64**	3.53
areific	Deuteranopia	12%	**37%**	8.73	**30.04**
imwaffe	Deuteranopia	9%	**21%**	4.88	**14.92**
gustav	Normal	4%	3%	0.35	0.48
troots	Normal	4%	3%	0.70	0.74

Tab. 2 - Table showing the error rates and scores for 6 out of the 16 subjects in the control group, for which the deficiency was established using both Ishihara Plates and Anomaloscope.

In Tab. 2 are shown the error percentages and the scores obtained along the two sets of possible confusion lines for the subjects in the control group; only 6 out of the 16 subjects are shown for means of readability. Errors and scores along the tritan lines are not shown since for the time being it has not been feasible to include tritan subjects in the control group, making it impossible to evaluate the effectiveness and accuracy of our data analysis. Error rates are calculated as the ratio between the number of errors committed by selecting a color belonging to the specific set of confusion lines over the total rounds the user played as shown in Eq. (2), while scores are computed as shown in Eq. (1). It can be seen that subjects manifesting a form of dyschromatopsia presents a higher error rate and score with respect to the normal trichromate subjects along both directions, while both dischromatoptics and anomalous trichromats shows only higher scores in both direction with respect to normal subjects, but similar error rates. Looking both at the error rates and the scores it can be seen that the maximum value coincides with the actual deficiency observed in the user. A player is marked as normal if at least one of the scores is less than 1 and deficient otherwise, with the specific direction determined by the maximum value among the two scores.
At the time of writing, data collection is still undergoing, and metrics might face changes and improvement in the near future, but the aim of this preliminary analysis is to ascertain whether variability in devices, viewing conditions and attention levels are by any means tolerable for objective data gathering, which seems the case given the high success rate of this preliminary assessment. This game is part of a research project started in 2014 (Eschbach, *et al.*, 2014; Rizzi, *et al.*, 2014) and enhanced in (Eschbach and Nussbaum, 2021) which aims at determining the role of the scene spatial arrangement in color vision, for CDOs and CNOs. The research project is still in course and the results of this color vision online game will be soon published.

Conclusion

Today, to screen and assess color blindness and color vision deficiency there are several clinical tests, which require qualified medical personnel, high concentration by the observer and often a preliminary observer education on the use of the instruments. These requirements are necessary to have a diagnosis with high rate of reliability but could be unsuitable for young children or people with behavioral problems. In this context, the development of web-based tests could be a solution to gamify classical clinical tests and collect large amounts of data, to make research, improve the current diagnosis methodologies and raise the awareness of common people to color vision issues. Clearly, an online test could never substitute a clinical test, but could be useful to sensitize the general public on color vision issues and to help researchers in collecting data and developing innovative color vision tests.

In this paper, we investigated the usage of a web-based game as a source for larger amounts of data related to color perception, providing some tips and tricks to design online color games and providing possible solutions. In addition, we also present a practical example: the game Qolour. This game has been specifically developed to study the role of borders and edges in color vision, but its implementation could be an example for any kind of research in color vision.

The main advantage of online games is the large amount of data which could be collected in a short time, and which could provide an overall averaging of the outcomes and a consequent increase of reliability and confidence during analysis. Furthermore, the user engagement in progressing the game provides data reliability, together with the possibility to involve in the study also children or young people.

References

Birch, J. (2012), ‘Worldwide prevalence of red-green color deficiency’, JOSA A, 29(3), pp. 313-320.

Cole, B. L. (2007), ‘Assessment of inherited colour vision defects in clinical practice’, *Clinical and experimental optometry*, 90(3), pp. 157 - 175.

EIZO (2022), Color Management for Creators - Mobile Devices [Online], www.eizoglobal.com/library/management/color-management-mobile-devices/ [20 July 2022].

Eschbach, R. *et al.* (2014), ‘Feeling edgy about color blindness’, *Electronic Imaging: Science and Technology*, SPIE, 2014(1), pp. 99 – 1.

Eschbach, R. and Nussbaum, P. (2021), ‘Examining spatial attributes of color deficient observers’, *Electronic Imaging: Science and Technology*, SPIE, 2021(16), pp. 308 – 1.

Hunt, D. M. and Carvalho, L. S. (2016), ‘The genetics of color vision and congenital color deficiencies’, *Human Color Vision*, Springer, pp. 1 - 32.

Judd, D. B., (1945), ‘Standard response functions for protanopic and deuteranopic vision’, *JOSA*, 35(3), pp. 199-221.

National Research Council (1981), *Procedures for testing color vision: report of working group 41*, Washington: National Academy Press.

Rizzi, A., *et al.* (2014), ‘Modified Ishihara test to study the role of edges in color discrimination’, *Italian Journal of Aerospace Medicine*, 11(2014).

Rodriguez-Carmona, *et al.* (2021), ‘Color vision assessment-2: Color assessment outcomes using single and multi-test protocols’, *Color Research & Application*, 46(1), pp. 21 - 32.

W3C, (2022), ‘CSS Color Module Level 3’ [Online], www.w3.org/TR/css-color-3/#rgb-color [22 July 2022].