QuickEval: A Web Application for Psychometric Scaling Experiments

Khai Van Ngo, Jehans Jr. Storvik, Christopher André Dokkeberg, Ivar Farup, and Marius Pedersen

Gjøvik University College, Gjøvik, Norway

ABSTRACT

QuickEval is a web application for carrying out psychometric scaling experiments. It offers the possibility of running controlled experiments in a laboratory, or large scale experiment over the web for people all over the world. It is a unique one of a kind web application, and it is a software needed in the image quality field. It is also, to the best of knowledge, the first software that supports the three most common scaling methods; paired comparison, rank order, and category judgement. It is also the first software to support rank order. Hopefully, a side effect of this newly created software is that it will lower the threshold to perform psychometric experiments, improve the quality of the experiments being carried out, make it easier to reproduce experiments, and increase research on image quality both in academia and industry. The web application is available at www.colourlab.no/quickeval.

Keywords: image quality, psychometrics, scaling, web, application, QuickEval

1. INTRODUCTION

The perceived quality of images is very important for the end-user, industry, engineers, and researchers. Subjective image quality assessment is considered the most precise way to quantify quality [1]. Assessment of image quality is a challenge for many in the field of image quality. The threshold for conducting subjective experiments can be high; they are time-consuming, they can be difficult to set-up and carry out, they might require special software, and so on. The most common methods for conducting such experiments include paired comparison, rank order, and category judgement [1]. In paired comparison experiments observers judge quality based on a comparison of image pairs, and the observer is asked which image in the pair is the best according to a given criterion, for example which has the highest quality. For rank order experiments the observer is presented with a number of images, who is asked to rank them based on a given criterion. Rank order can be compared to doing a pair comparison of all images simultaneously. In category judgment the observer is instructed to judge an image according to a criterion, and the image is assigned to a category.

These types of experiments are also very often carried out in controlled laboratory settings with a limited group of observers. These often rely on software that has been designed to cover the functionality required for that specific experiment, and the software is very often not publicly available. Attempts have also been made to allow web-based image quality experiments. Qiu and Kheiri [2] presented Social Image Quality a web-based system for paired comparison experiments. On their website the user is shown two images, and is asked which image has the highest quality. The displayed images are decided by the webmaster of the webpage, and does not allow researcher to upload images. In the TID2008 [3] and TID2013 [4; 5] databases, experiments were also conducted online in addition to laboratory experiments. Analysis of the results showed little difference between controlled and uncontrolled experiments [3]. In the uncontrolled experiments by Simone et al. [6] a website was designed based on category judgement specifically for that experiment. Analysis of the results showed also no significant difference between the uncontrolled experiment and a controlled experiment carried out in a lab. These findings support that image quality experiments could be carried out through the web.

The number of observers in subjective experiments is important for the statistical analysis usually performed on the results, and the precision of the statistics [1; 7]. When a large number of observers are used, the average is more likely to be consistent with overall image quality [8] and the precision of the estimated values

---

Further author information: Send correspondence to M.P. E-mail: marius.pedersen@hig.no

increases [1]. A number between 10 and 30 observers are recommended for typical scaling applications [1; 7]. There is usually a trade-off between the number of stimuli and the number of observers. Due to time restrictions it is often more desirable to have a large number of observers than a large stimuli material [9]. When a large number of observers is required, a software specifically designed for the purpose will ease the work of conducting psychometric experiments.

To the best of our knowledge a tool for conducting paired comparison, rank order, and category judgement does not exist in a single software. We have also not been able to find software allowing for rank order experiments. In this paper we propose a web application, QuickEval, to create and carry out psychometric experiments both for controlled and uncontrolled environments. The type of psychometric experiments supported by the software is paired comparison, category judgement and rank order. It also supports experiment on a large scale, enabling experiments over the Internet.

The paper is organized as follows: first we present relevant background, then we introduce the proposed application, at last we conclude and propose future work.

2. BACKGROUND

2.1 Paired comparison

In paired comparison the observer is shown a pair of stimuli, with or without an original, they are asked which of the stimuli are is the best according to a criterion, such as which has the highest quality or which is the most pleasant. These experiments can be forced-choice, i.e. that the observer needs to make a decision, or not forced-choice, i.e. the observer does not need to make a distinction between them (a tie). Observers judge n reproductions for m original stimuli, which results in $(n(n-1) \times m)/2$ comparisons. Usually, each pair is shown twice, where the left and right position has changed, giving $mn(n-1)$ comparisons. Paired comparison is a popular model for image quality experiments, since it is simple and require little knowledge by the user. It has been often used for evaluation of gamut mapping [10–13].

The collected data is usually transformed into interval scale data following Thurstone's Law of Comparative Judgement [14]. The transformation results in the distance of a given stimulus from the mean score of the data set.

The selected stimulus from a pair is recorded in a $n \times n$ raw data matrix. One is given in the column i and row j for stimulus i compared to stimulus j. Summing all observer raw matrices results in a summed $n \times n$ frequency matrix. A percentage matrix is computed by dividing by the number of observations. A Logistic Function Matrix (LFM) is computed from the summed frequency matrix by using the formula from Bartleson [15]:

$$LFM = \ln\left(\frac{f+c}{N-f+c}\right),\tag{1}$$

where f is the value from the summed frequency matrix, N is the number of observations, and c is an arbitrary additive constant. A c of 0.5 is the most commonly used [15–17].

A scaling coefficient is used to transform the LFM into z-scores. The coefficient is calculated by taking the linear regression between the inverse of the cumulative standard normal distribution for the percentage matrix and the LFM values, where the coefficient is the slope of the regression line. The z-score matrix is found by multiplying the coefficient with the LFM. Further, mean values for the reproductions can be found by averaging.

95% confidence intervals (CIs) are usually reported together with the z-scores. The most common method to calculate the 95% CI is based on the number of observations (N) and the standard deviation (σ).

$$CI = 1.96 \frac{\sigma}{\sqrt{N}},\tag{2}$$

Z-scores have a scale with units equal to $\sigma\sqrt{2}$, and therefore σ can be set to 1 [17].

2.2 Rank order

In rank order experiments the observer is presented with a set of images, and is asked to rank them according to a criterion, for example from highest to lowest quality. Rank order can be considered as doing multiple paired comparisons at the same time. If the set contains many images the task can become complex for the user, but it is often considered as a fast and simple experimental method to judge images. Rank order has often been preferred when evaluating gamut mapping algorithms on paper [11; 18].

To analyze the results from rank order the rank of each stimulus is collected. Using comparative paired comparison modeling the collected data can be converted into paired comparison data [19]. The transformation into a paired comparison matrix where stimulus i gives a scaled value (s_i):

$$s_i = \frac{1}{(n-1)} \sum_{j=1}^{n} z \left(\frac{\sum_{j=1}^{N} p_{kij}}{N} \right), \tag{3}$$

where $z()$ represent the operator to convert the proportion of choice to z-scores as in paired comparison, p_{kij} is a binary number (1 represent that stimulus i is greater than stimulus j, and 0 vice versa, for observation k), n is the number of stimuli, and N is the number of observers.

$$RS_i = \frac{1}{(n-1)N} \sum_{k=1}^{N} \sum_{j=1}^{n} p_{kij} = \frac{1}{n-1} \sum_{j=1}^{n} \left[\frac{\sum_{j=1}^{N} p_{kij}}{N} \right], \tag{4}$$

The $z()$ operator is applied to to Equation 4 the scaled ranked (SR_i) values are obtained:

$$SR_i = z \left(\frac{1}{n-1} \sum_{j=1}^{n} \left[\frac{\sum_{j=1}^{N} p_{kij}}{N} \right] \right). \tag{5}$$

2.3 Category judgement

In category judgement the observer is asked to judge an image according to a criterion, and then assign it to a category. It is most common to use five or seven categories, allowing for a mid-point. An advantage of this method over rank order and paired comparison is that the distance between images is recorded. Category judgement is considered to be more complex than the two other methods, but it is rather fast since it requires fewer comparisons than paired comparison. It is usually the preferred method when using a large dataset of images, such as the Colourlab Image Database: Image Quality (CID:IQ) [20].

Torgersons Law of Categorical Judgment [21] is one way to analyze ordinal data. A procedure for analyzing data from category judgment experiments is given by Engledrum [1] .

A $n \times (m+1)$ raw matrix is computed, where n is the number of stimuli and m the number of categories. In this matrix each column contains the frequency that one stimulus is judged to belong to a specific category. Further, this matrix is summed in order to obtain the frequency matrix K, being the basis for a cumulative frequency matrix. An $n \times (m+1)$ cumulative percentage matrix is calculated by first determining the cumulative matrix CM:

$$CM = \sum_{k=1}^{g} = K_{jk}, \tag{6}$$

where $g = 1, 2, ..., m+1$, and g is the number of categories, and j represent the stimuli.

The next step consists of transforming the cumulative frequency matrix into a cumulative percentage matrix (PM)

$$PM = \frac{CM_{jg}}{CM_{j,m+1}}. \tag{7}$$

Values in the last column of PM are 1, and are dropped so that P becomes $n \times m$. This results in the boundaries of the first and last categories are open-ended intervals and can not be determined [1].

Z-scores are calculated by applying the inverse of the standard normal cumulative distribution function to PM. The final scale values (z-scores) for each stimulus are determined by subtracting the average of each row from the mean of the matrix of the preliminary z-score matrix [22]

$$S_j = \frac{1}{mn} \sum_j \sum_g z_{j,g}. \tag{8}$$

3. QUICKEVAL - THE APPLICATION

QuickEval is a web application where the researcher can set up experiments, invite observers to carry out the experiment, and extract the observer results easily for the methods paired comparison, rank order, and category judgement. The application supports two different modes; observer mode and scientist mode, where scientist mode has functionality and options relating to creating and managing experiments, and the observer mode have features related to performing experiments created by scientists. QuickEval also follows best practice and common guidelines in subjective image quality evaluation [10]. The web application is freely available at `www.colourlab.no/quickeval`. Figure 1 shows the front page of the application, where observers can select experiments.

3.1 Scientist mode

The scientist mode gives the scientist functionality like managing pictures and picture sets, managing experiments, as well as to analyze and treat result data from experiments. Other important functionality is the ability to export either raw or calculated data like z-scores with graphs and tables. Figure 2 shows the scientist dashboard.

The work-flow of creating an experiment consist of the scientist uploading picture set(s), choosing the experiment parameters and options like method, name, whether it is to be public, the order of picture set(s) and instructions. Other options worth mentioning are random unique picture queues for every observer (the scientist can of course make their own custom picture queue as well), set background color for the experiment and whether or not to show the time spent by an observer. The scientist can also choose to set the experiment as public - being available to everyone, or hidden - being only available to invited observers.

3.2 Admin mode

The admin mode extends the scientist mode and provides administrator privileges in order to manage users, experiments, and images. The most important functionality in the admin mode is the option of granting scientist privileges to a registered observer, as you cannot register as a scientist in QuickEval right away. Administrators can also delete images, experiments and anonymous users that are older than a given date. It is created in order to have an easy way to keep maintain QuickEval without having the hassle of running SQL queries through the database, or deleting images manually. Its features also keeps the redundancy intact of the database by removing all the correct relations and database posts.

3.3 Observer mode

A first time observer can opt to register as an observer in the observer mode with name, age, country, affiliations etc., or they can choose to perform experiments as anonymous, although some experiments might require the observer to give information such as age or country. This is determined by the scientist. When logged in, observers is presented with a list of all public experiments from every scientist registered with QuickEval as shown in Figure 1. Experiments may be set as hidden if the scientist doesn't want anyone to carry out their experiment, or if only invited people are allowed to take the experiment. Some experiments may not be unlocked before the observer has undergone a simple Ishihara test to determine whether the observer is to some extent color deficient or not.

Figure 1. Front page and search of experiments. Observers can select an experiment from the list of experiment, and then get a description of the experiment, for starting the experiment they simply click "Start Experiment". In some experiments the scientist can decide to ask for information about the observer, in this example the firstname and age is required to start the experiment.
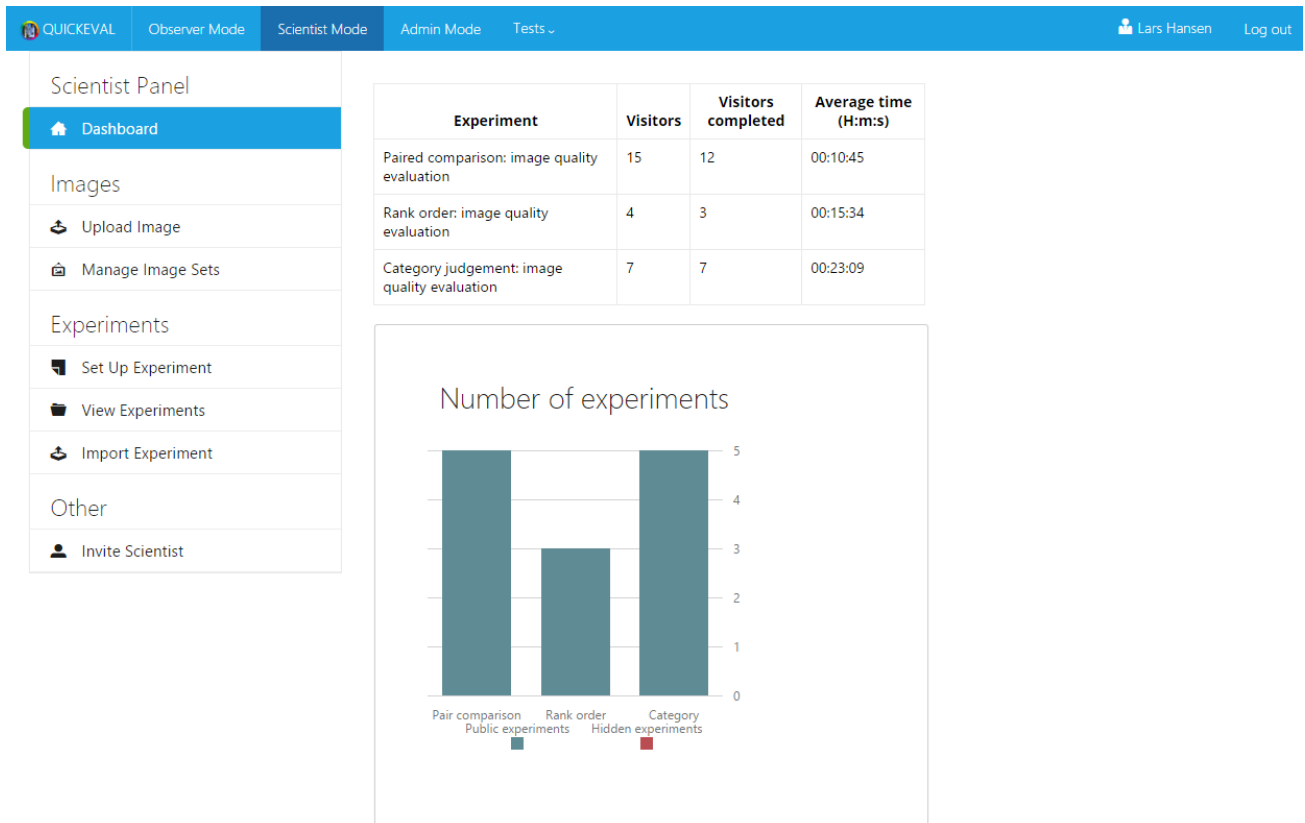
Figure 2. Scientist dashboard. This is where the scientist gets information about his experiment, such as the number of visitors and how many that have completed the experiment. The scientist can also manage image sets, set up experiments, view on-going experiments, or expert experiments.

Experiments can be sorted by method (rank order, paired comparison, and category judgement), institution, organization or name of the scientist. Searching for a given experiment is also possible through the use of the search bar. Once the wanted experiment has been selected, it can be started if all prerequisites have been met. This could be name, age and so on and is something the scientist sets up at creation of the experiment.
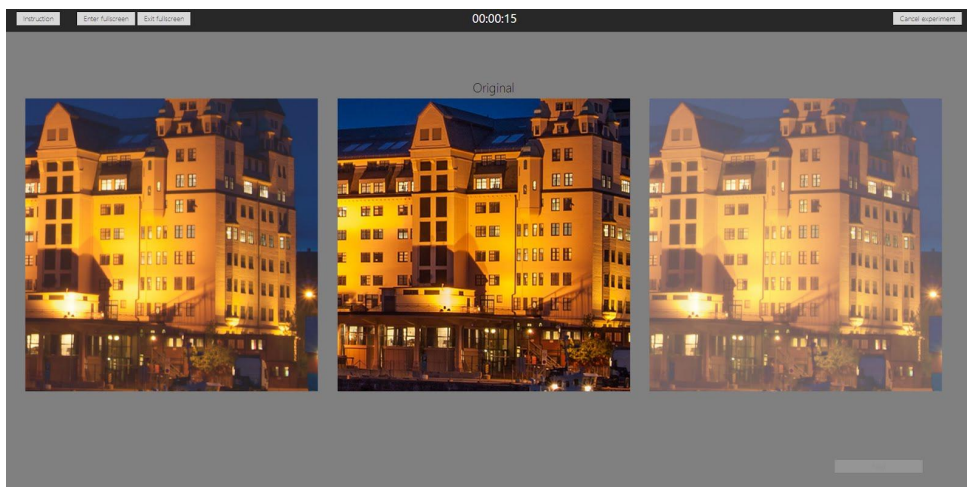

Figure 3. A common paired comparison experiment is shown. The original image is shown in the middle with two different reproductions on each side.
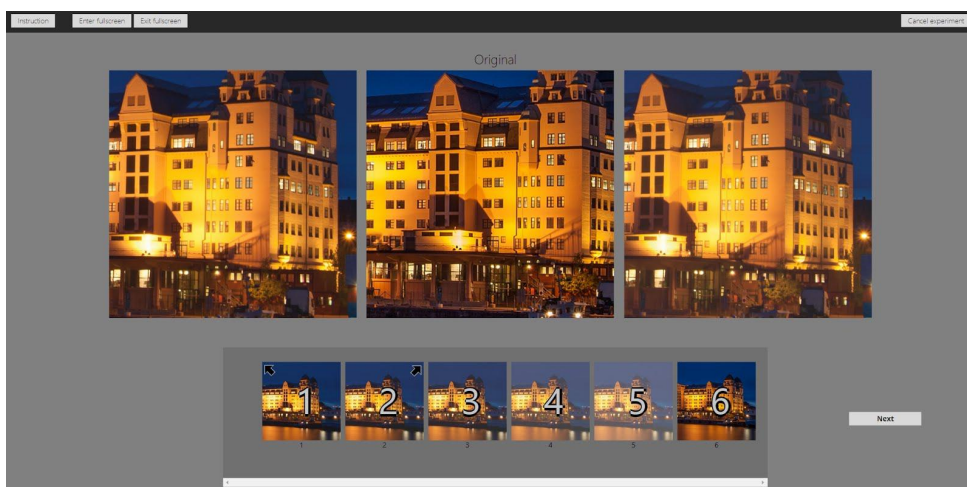

Figure 4. A common rank order experiment is shown. The observer is in this example asked to rank 6 different reproductions. He/she can choose which reproductions to compare against the original (shown in the middle). The arrows (left/right) indicates which image is shown to the left or to the right. The observer simply ranks the reproduction by drag and drop.

3.4 Execution

Once an experiment has been started, the observer is taken to a screen with the pictures that are to be evaluated and belonging instructions. This screen only contains the actual pictures. Great care has been taken in order not to put any other disturbing elements like unnecessary buttons and intense colors. The environment of the experiment execution consist of only the out most necessary functions, and have by default only elements in neutral grey and white, and can be further customized by the scientist through experiment parameters. The picture arrangement shown differs dependent on the method of the experiment, and the picture queue.

Paired comparison experiment will show only two pictures on either side with the option to have original in

the middle as shown in Figure 3. The observer simply left clicks on the one that is the best one accordingly to the given instructions, or uses the left or right arrow keys, and in some experiments the the user can also decide that the reproduction are identical and rate them as equal if the scientist allows them. The picture is then highlighted and the observer can proceed to the next step in the experiment.

Rank order will show all the pictures in the set on screen like shown in Figure 4. In order to properly evaluate every picture, the user has the possibility to drag them on screen two at the time in order to compare them to each other. An arrow is used on the picture-list below to indicate which pictures are being evaluated at the screen, as this would be the least distracting option compared. The pictures are rated with drag and drop, and numbers on the thumbnails are used in order to keep track of the pictures when ordering. Once all pictures have been compared, the user can move on to the next picture set.

Category judgement will show only the reproduction with the option of having the original picture on the left side as shown in Figure 5. A drop-down menu is placed beneath the reproduction with the categories set by the scientist for the experiment, and the user simply picks what category fits best. The drop-down menu default is set to "Choose a category". Once the category has been chosen, the user can move on to the next picture.

Pictures that are larger than the view area has the ability to be panned. All pictures will then be panned synchronized so that the observer is able to compare the same area across the reproductions and the original simultaneously. To pan the observer have to click and hold on the either picture and move around. Results from the carried out experiments are stored in a database for later use and analysis by the scientist.
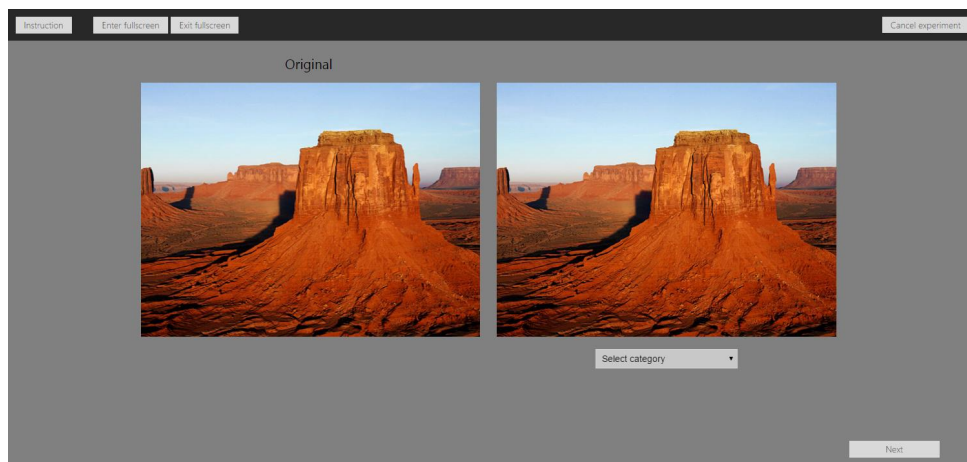

Figure 5. A common category judgement experiment is shown. The observer is in this example asked to choose category for different reproductions.

3.5 Calculation and visualization of results

QuickEval is able to view the results directly in the application as dynamic and interactive graphs, tables with z-scores shown in Figure 6. Graphs can be exported as either .svg, .png, .jpg or .pdf format to be used in for example presentations and articles. In addition to the graphs tabulated values with the mean z-score and upper and lower limits for the 95% confidence interval is displayed. The z-score features is the latest module to be added to QuickEval and offers an streamline work-flow for the scientists.

3.6 Exporting results and experiments

The results from an experiment can be exported by the scientist in a raw .csv-format shown in Figure 7 with the ability to choose which attributes are to be exported. The export function enables the researcher to easily import the raw data into other tools for post-processing.

QuickEval also offers the option of exporting entire experiments with its resources, including the results, image sets and experiment parameters. This enables the scientist to download a local copy of the experiment, either for backup or for conducting further experiments at a later time. The export of entire experiments is done
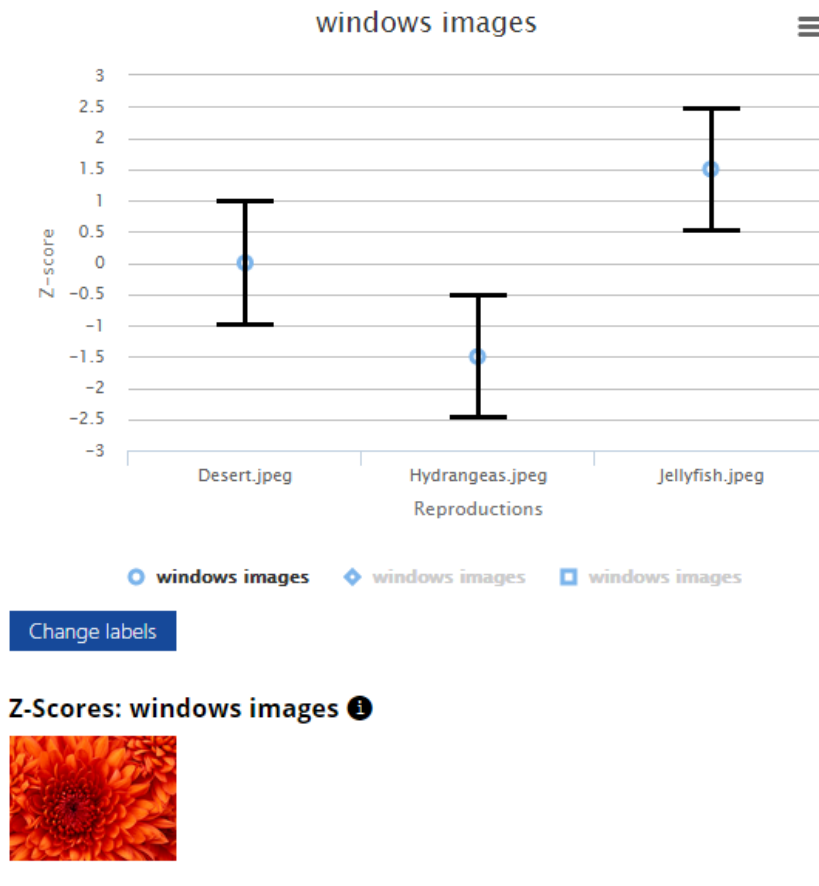
Figure 6. Z-Score plot.

in JSON format, and the scientist will get all files in a zip file. The exported experiments can easily be imported in QuickEval.

3.7 Implementation details

QuickEval as a web application utilizes and relies heavily on JavaScript with the addition of jQuery and Ajax in the front-end, and uses PHP- and MySQL-based back-end. This makes it possible to provide a fast responding and feature rich application. Most of the work done by the application is done client side which lightens the load on the server giving the user a user experience that is fast even if there is a larger workload and many users. A client heavy web-application also seems more interactive than most web pages, as it removes most of the need to reload the website while navigating.

The design of the application is based on what is considered as modern and up to date, that means a flat, squared and minimalistic design. Inspired by services and products like Facebook, Android, and Windows 8,

```
 1  ***Metadata***
 2  Name: Rank order: image quality evaluation
 3  Description: Rank order: image quality evaluation
 4  Type of experiment: Rank order
 5  Timestamp experiment creation: 2014-11-26 18:52:29
 6  Timestamp export: 2014-11-29 10:00:41
 7  Scientist: 1
 8  Number of observers: 2
 9  Number of instructions: 0
10  Background colour: backGroundcolor
11  Viewing distance: 1
12  White point:
13  Screen luminance:
14  Ambient illumination:
15
16  ***Parameters***
17  Allow colour deficiency: NO
18  Display timer: NO
19  Display original: NO
20  Hidden experiment: YES
21
22  ***Instructions***
23
24  ***Image Set***
25  Imageset1,windows images,Desert.jpeg,Hydrangeas.jpeg,Jellyfish.jpeg
26  Imageset2,windows images,Desert.jpeg,Hydrangeas.jpeg,Jellyfish.jpeg
27  Imageset3,windows images,Desert.jpeg,Hydrangeas.jpeg,Jellyfish.jpeg
28
29  ***Observer data***
30  Observer1,Lars Hansen,lars.hansen@gmail.com,Norway,male,45,1,Superuser,Windows,1920,955,2014-11-26
31  Observer2,Anonymous21 ,,,,,,1,Anonymous,Windows,1920,955,2014-11-26 18:53:16,2014-11-26 18:53:16
32
33  ***Input fields***
34
35  ***Input field results***
36
37  ***Experiment results***
38  Imageset1,Lars Hansen,2014-11-26 18:52:45,1,2,3
39  Imageset1,Anonymous21 ,2014-11-26 18:53:07,3,1,2
40  Imageset2,Lars Hansen,2014-11-26 18:52:47,1,2,3
41  Imageset2,Anonymous21 ,2014-11-26 18:53:10,1,3,2
42  Imageset3,Lars Hansen,2014-11-26 18:52:50,1,2,3
43  Imageset3,Anonymous21 ,2014-11-26 18:53:17,2,1,3
```

Figure 7. Example of exported data from an experiment. The exported data contains both meta-data about the experiment, information about the image set, and the recorded data.

the result is a design that utilizes elements that is similar to Windows 8. The Metro UI CSS 2.0 by Sergey Pimenov [23] provides the needed styles and elements to create a minimalistic and intuitive interface that convey information effectively.

To support various features like picture panning, navigation and uploading of pictures, QuickEval uses plugins and libraries that are licensed under open source. They provide the application with a wide support when it comes to different ways a experiment might be set up and future proofing in terms of scalability of users and features.

The technology running the database is as mentioned earlier MySQL, and is designed as a relations database. The technology provide support for foreign keys and dependencies between tables and attributes which is required to support the all the features. By supporting a wide array of features and parameters for each type of experiment and users, the database have to be able to scale accordingly to today's and future needs. With that being said,

the database is designed in a way that is able to provide QuickEval with future support for additional experiment types and parameters for experiments. Hard facts about the database; it consist of 20 tables, where each contains anything from three to 23 attributes, these tables are then connected together with 33 relations.

3.8 Performing large scale controlled and uncontrolled experiments

Image quality assessment experiments are often performed in a laboratory. This sometimes limits the type of experiments to invite only. QuickEval allows you to perform both controlled and uncontrolled experiments - and unlike manual experiments in laboratories, requires less effort in terms of time managing, scheduling and post-processing.

The scientist is allowed to have his or her experiments fully open or make them invite-only. Open experiments allows the scientist to create an experiment, wait for results, then have the results calculated into a more comprehensive format. By using other sources like social media or the like, the scientist can share the link effortlessly and reach a great number of potential observers. Everyone visiting the web page will also be able to browse all the experiments and choose which one to partake in.

Invite-only experiments also shares some of the same benefits as open experiments. The scientist can share the link for the closed experiments through his own circles or organization to target a specific audience. Another use of closed experiments is to perform them in a laboratory, where everything from monitors and lighting is controlled - very similar to performing manual experiments - but letting the application handle both the set-up, execution and calculations.

3.9 Applications

QuickEval is already in use at Gjøvik University College for teaching psychometrics, and has been welcomed by the students. It has also been used in research, Le Moan et al. [24] used QuickEval to create a database for spectral image quality.

4. CONCLUSION AND FUTURE WORK

QuickEval is a unique web application for carrying out psychometric scaling experiments, such as paired comparison, rank order, and category judgement. It allows to set-up and manage experiments both in a laboratory and over the internet. It is also the first software to support rank order. QuickEval offers a simple and effective way to set-up, conduct, and analyze psychometric scaling experiments. The web application is available at `www.colourlab.no/quickeval`.

Future work include extending the functionality of QuickEval, such as supporting confidence intervals from Montag [25] and calculation of mean opinion scores. Future work can also include additional statistical tests on the experiment results. We would also like to conduct large scale testing of the application in order to improve it further.

5. ACKNOWLEDGEMENTS

References

[1] Engeldrum, P. G., [*Psychometric Scaling, a toolkit for imaging systems development*], Imcotek Press Winchester USA (2000).

[2] Qiu, G. and Kheiri, A., "Social image quality," in [*Image Quality and System Performance VIII*], Farnand, S. P. and Gaykema, F., eds., *Proceedings of SPIE* **7867**, 78670S (Jan 2011).

[3] Ponomarenko, N., Lukin, V., Zelensky, A., Egiazarian, K., Astola, J., Carli, M., and Battisti, F., "Tid2008 a database for evaluation of fullreference visual quality assessment metrics," *Advances of Modern Radioelectronics* **10**, 30–45 (2009).

[4] Ponomarenko, N., Ieremeiev, O., Lukin, V., Egiazarian, K., Jin, L., Astola, J., Vozel, B., Chehdi, K., Carli, M., Battisti, F., and Kuo, C.-C. J., "A new color image database tid2013: Innovations and results," in [*Proceedings of ACIVS*], 402–413 (Oct 2014).

[5] Ponomarenko, N., Ieremeiev, O., Lukin, V., Egiazarian, K., Jin, L., Astola, J., Vozel, B., Chehdi, K., Carli, M., Battisti, F., and Kuo, C.-C. J., "Color image database tid2013: Peculiarities and preliminary results," in [*4th Europian Workshop on Visual Information Processing*], 6p. (Jun 2013).

[6] Simone, G., Pedersen, M., and Hardeberg, J. Y., "Measuring perceptual contrast in uncontrolled environments," in [*European Workshop on Visual Information Processing (EUVIP)*], 102–107, IEEE, Paris, France (Jul 2010).

[7] Engeldrum, P. G., "Psychometric scaling: Avoiding the pitfalls and hazards," in [*Image Processing, Image Quality, Image Capture Systems Conference*], 101–107, IS&T, Montréal, Quebec, Canada (Apr 2001).

[8] Lundström, C., "Measuring digital image quality," tech. rep., Center for medical image science and visualization, Linköping University (2006). 04/06/09: `http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-5586`.

[9] Sharma, G., [*Digital Color Imaging Handbook*], CRC Press, Inc., Boca Raton, FL, USA (2002).

[10] CIE, "Guidelines for the evaluation of gamut mapping algorithms," Tech. Rep. ISBN: 3-901-906-26-6, CIE TC8-03 (156:2004).

[11] Dugay, F., Farup, I., and Hardeberg, J. Y., "Perceptual evaluation of color gamut mapping algorithms," *Color Research & Application* **33**, 470–476 (Dec 2008).

[12] Bando, E., Hardeberg, J. Y., and Connah, D., "Can gamut mapping quality be predicted by color image difference formulae?," in [*Human Vision and Electronic Imaging X*], Rogowitz, B., Pappas, T. N., and Daly, S., eds., *SPIE Proceedings* **5666**, 180 – 191 (Jan 2005).

[13] Bonnier, N., Schmitt, F., Brettel, H., and Berche, S., "Evaluation of spatial gamut mapping algorithms," in [*Color Imaging Conference*], 56–61, IS&T/SID, Scottsdale, AZ (Nov 2006).

[14] Thurstone, L. L., "A law of comparative judgment," *Psychological Review* **34**, 237 (1927).

[15] Bartleson, C. J., [*Optical Radiation Measurement. Vol. 5 - Visual Measurements*], Academic Press Inc. (1984).

[16] Green, P. J. and MacDonald, L., eds., [*Colour Engineering: Achieving Device Independent Colour*], John Wiley & Sons (2002).

[17] Morovic, J., *To Develop a Universal Gamut Mapping Algorithm*, PhD thesis, University of Derby (1998).

[18] Bonnier, N., Schmitt, F., Hull, M., and Leynadier, C., "Spatial and color adaptive gamut mapping: A mathematical framework and two new algorithms," in [*Color Imaging Conference*], 267–272, IS&T, Albuquerque, NM (Nov 2007).

[19] Cui, C., "Comparison of two psychophysical methods for image color quality measurements: Paired comparision and rank order," in [*Color Imaging Conference*], 222–227, IS&T/SID, Scottsdale, AZ (Nov 2000).

[20] Liu, X., Pedersen, M., and Hardeberg, J., "Cid:iq - a new image quality database," in [*Image and Signal Processing*], Elmoataz, A., Lezoray, O., Nouboud, F., and Mammass, D., eds., *Lecture Notes in Computer Science* **8509**, 193–202, Springer International Publishing, Cherbourg, France (Jul. 2014).

[21] Torgerson, W. S., [*A Law of Categorical Judgment*], ch. Consumer Behaviour, 92–93, New York University Press (1954).

[22] Green, P. J., *Gamut mapping and appearance models in colour management*, PhD thesis, University of Derby, UK. (2003).

[23] Sergey S., P., "Metro UI CSS : Metro bootstrap CSS library." `http://metroui.org.ua/`, visited: 30/11/14.

[24] Moan, S. L., George, S. T., Pedersen, M., Blahova, J., and Hardeberg, J. Y., "A database for spectral image quality," in [*Image Quality and System Performance XI*], Larabi, M.-C. and Triantaphillidou, S., eds., **9396**, 9396–25 (Feb. 2015).

[25] Montag, E. D., "Empirical formula for creating error bars for the method of paired comparison," *Journal of Electronic Imaging* **15**, 222–230 (Jan-Mar 2006).