Methods in Psychology 5 (2021) 100077

Contents lists available at ScienceDirect

Methods in Psychology

journal homepage: www.sciencedirect.com/journal/methods-in-psychology

Psychophysical scaling method for measurement of colors concept in children and adults

Marcelo Fernandes Costa [a,b], Carlo Martins Gaddi [b,*], Vitor Melo Gonsalez [a], Fraulein Vidigal de Paula [c]

[a] *Dept. Psicologia Experimental, Instituto de Psicologia, Universidade de São Paulo, São Paulo, São Paulo, Brazil*
[b] *Núcleo de Neurociências e Comportamento e Neurociências Aplicada, Universidade de São Paulo, São Paulo, São Paulo, Brazil*
[c] *Dept. Psicologia do Aprendizado, da Educação e da Personalidade, Universidade de São Paulo, São Paulo, São Paulo, Brazil*

A R T I C L E I N F O

Keywords:
Color salience
Law of comparative judgment
Color concept
Psychophysical scaling
Children

A B S T R A C T

Our study aimed to present a new method to generate a scale for color concept distances using the Law of Comparative Judgment and derivate an interval scale for the amount of color concept salience. A hundred thirty-eight participants (mean age = 22.7 yrs; SD = 1.9) with normal visual acuity and color vision were evaluated. Additional 225 children were also evaluated (mean age = 8,9yrs; SD = 2.3). The task consisted of writing a list of the one-term word for colors and data were analyzed based on Thurstone's rank order scaling method. For adults, the most salient concept was blue and the less was violet with a distance for concepts of 3.51 units between them, whereas for children the most salient concept was blue and the less was violet with a distance for concepts of 3.67 units between them. Significant differences between adults and children occurred for white, brown, pink, orange, gray, and violet. The internal consistency of our measurement was lower than 0.1% for almost all colors which suggests high reliability for both age groups, except for brown and anil in children. Thus, our innovative approach produces quantitative interval class data rather than categorical or ordinal data allowing quantitative measurement of the amount of the concept salience on a one-dimensional scale, feasible to be used for adults and children.

1. Introduction

Color is a visual dimension widely used to study basic aspects of cognitive functions as categorization, recognition, and concept formation. It is demonstrated that color concept is effortful to be learned by children based on the discrepancy in their ability to learn the names for everyday objects and the tardy, erratic nature by which they learn color terms (Pitchford and Mullen, 2001) and it is strongly dependent of the integrative conditions like maternal input, the individual color preferences, and perception (Pitchford and Mullen, 2005). The majority of the literature of the color concept studies, the color naming tasks, and color categorization tasks (Buchsbaum and Bloch, 2002; Cole et al., 2006; Rosch, 1973; Taft and Sivik, 1997; M. A. Webster and Kay, 2012; Zollinger, 1988), used color samples in objects or on monitor screens to evaluate color concepts.

Athanasopoulos et al., (Athanasopoulos et al., 2011), studying the cognitive representation of color concepts in bilingual Japanese-English, found that Japanese monolingual judged two perceptual colors to be less similar if they fell into different linguistic categories in Japanese than if the two colors were from the same category. English monolingual tested with identical stimuli did not show a categorical perception under any condition tested by authors. Japanese–English bilingual displayed a cognitive pattern that was "in-between" the two monolingual groups, suggesting that knowledge of two languages with contrasting ways of parsing reality has profound consequences for cognition. Based on those findings, we consider that language (in this case, names or terms) is a variable determinant to concept formation. For that reason, an experimental approach that uses the only language as a variable would address a deeper relationship that establishes concept formation, the relationship between the abstract representation and the representation of language (i. e., between concept and their terms for colors). Such an approach would additionally avoid those existing experimental problems with the association between the linguistic with the object color, selection of color samples, and others.

* Corresponding author. Marcelo Fernandes Costa, Universidade de São Paulo, Instituto de Psicologia, Departamento de Psicologia Experimental, Av. Prof. Mello Moraes 1721, Butantã, São Paulo, SP, 05508-030, Brazil.

E-mail address: costamf@usp.br (C.M. Gaddi).

https://doi.org/10.1016/j.metip.2021.100077
Received 21 May 2019; Received in revised form 13 July 2021; Accepted 7 September 2021
Available online 14 September 2021

Besides the tasks of naming and categorizing, a method more similar to the one we applied in our study is also used for measures of what is called color salience. This method has a task that allows the evocation of semantic responses without the need to use physical stimuli. This method is known as the Method of Listing (de Sousa et al., 2016; Sutrop, 2001; Thompson and Juan, 2006; Uuskula and Bimler, 2016). In that method, the participants make a free listing on a specific topic, assuming that lists reflect a subset of the total knowledge of the participant, and the most frequent terms are listed first, resulting in the first items listed are assumed the most culturally important. There are several ways of analyzing the outcome of a free listing task.

The Cognitive Salience Index analysis uses the term frequency to calculate the mean position parameter, which shows the psychological salience of a term into one quantitative parameter (Sutrop, 2001). This index performs a numerical measurement, but of the positional domain on the psychological continuum, supported by the central tendency of greater frequency, and not considering, the frequencies in the adjacent ranks, making it difficult to compare with the other frequencies. The CSI is calculated based on the following equation:

$$S = \frac{F}{N\ mP} \qquad \text{(Eq. 1)}$$

since the frequency (F) with which a term is named in the list task is divided by considers the weight of the mean position (mP) in which the term is named, and N is the number of subjects. Thus, considering that all subjects have named a hypothetical term, then ($F = N$), the mean position of that term will be 1 and the salience (S) is also 1 for that term. In other words, the Cognitive Salience Index could be considered numerical but in an ordinal measurement dimension, in which the salience is measured by its location (ordinal position) on a range between 0 and 1. In essence, this procedure determines the same class of results as those obtained in a location procedure of a given point on a finite line, a measurement at an ordinal level. In being an ordinal measurement, the distances between the numbers (labels) on that continuum are virtually non-existent, because the numerical difference measured doesn't express any ratio of a unit (distance), only the relative position in a scale (order). Thus, the Cognitive Salience Index outcome is a position determined on a dimensional range of the psychological continuum throttled between 0 and 1.

Although there was an advance in the representation of cognitive saliences with the Cognitive Salience Index, there was still one condition wasn't solved: the low information generated by the indices of the existing scales. An important advance would be the ability to evolve our measure to a dimension of greater quantitative representation, such as that found in interval measures. Also, if we considered the quantitative approach of the Cognitive Salience Index we still face the problem of determining the mean point without considering the spreading of the term on the list, reducing its accuracy (Smith and Borgatti, 1998; Smith et al., 1995).

In an attempt to obtain measurements as the "numerical amount" of a concept's salience, an interval or rational quantitative approach is necessary. In interval scales, the distances between each point are defined with high mathematical precision and they are, indeed, the first quantitative form in the ordinary sense of the word (Stevens, 1946). In comparison with the nominal or ordinal scales, we change from the determination of equality or order, respectively, as the basic empirical operations toward the determination of equality of intervals or equality of differences.

Interval scales can be built by using several procedures as equisection, bisection, and pair comparison. The latter was the procedure used by L. L. Thurstone to develop the Law of Comparative Judgment, that applies to comparisons of physical stimulus intensities, as well as to qualitative comparative judgments as those more frequently presented in educational, sociological, and psychological issues (Thurstone, 1927a, 1927b, 1929, 1931a).

The method consists in simplifying Fechner's constant method extracting, from the absolute rank order, proportions of judgments, i.e. pairs of stimuli that follow "A is greater than B", for every possible pair. Knowing the proportion of subjects who placed A and B in point x of n rank orders, we can calculate the proportional perception of $A > B$ by evaluating their distribution in the ranks for the entire group of subjects. More details about the procedures can be read in section 2.3 Psychophysical Scaling.

This procedure developed by Thurstone is in a way equivalent to the Cognitive Salience Index. However, we considered that all of those proportions of each sample at each rank position could, then, be used to generate a proportion matrix allowing the Case V method to calculate their respective interval distances (Thurstone, 1928). This means that our approach takes into account the level of precision of the quantified term, which represents an important advance in the measurement of color terms. There are clear advantages to using Thurstone's methodology, presented by Maydeu-Olivares et al., (Maydeu-Olivares and Böckenholt, 2008). For our study, the main advantages are the use of a simple list of words, which is a very simple procedure to be performed by any minimally literate subject; and that the validity of the inferences can be verified by measures of internal consistency, guaranteeing robust methodological and statistical significance. These advantages justify the choice of this method as preferred for subjective measures in health (Krabbe, 2008); The rank order procedure was successfully used to analyze specimens of handwriting with similar accuracy by association with Thurstone's Law of Comparative Judgment (Thurstone, 1931b) as those achieved obtained in other studies for health quality outcome (Krabbe, 2008; Maydeu-Olivares and Böckenholt, 2008), group support (Li et al., 2001), healthy states (Arons, Krabbe, Scholzel-Dorenbos, van der Wilt and Rikkert, 2012), discriminatory judgments of nationality or race preferences (Thurstone, 1928) and in measuring attitudes (Costa et al., 2014; Likert et al., 1993) which used the Thurstone's method. The step changed from nominal or ordinal to interval scale is huge and significant if we want to address questions like "What is the color with more amount of salience for the Brazilians? Has that color the same amount of salience for Americans? Or Norwegians?". Scholars working in semantics discuss salience in terms of accessibility and have suggested that it could be a result of high frequency (e.g. Bardovi-Harlig, 1987; Chapman and Tunmer, 1995; Giora, 2003; Cieślicka, 2006). They argue that more frequent features are more regular, making them more easily accessible and salient. Thus, the study of salience of color terms help us better understand different cultures.

Considering the possibility of applying this method in children, we evaluated the possibility of studying concepts of color using this methodology (Norman and Rieber, 1968). We found only one study on the concept of colors in children. This study showed that color concepts are an important cognitive activity for concept learning and formation. Those results are in line with other studies suggesting that color is an important attribute to object concept learning (Feher and Meyer, 1992).

The main purpose of this paper is to address the color concept trying to avoid the classical but less informative categorization or naming tasks, since our outcome intends to measure the color concept *continuum* in a quantitative interval scale, allowing the use of statistical tools instead of a fractioning (or frequency) presentation of the concepts on the ordinal scale. An important fact about scales in psychology is that they must be a representation of psychological continuums, and therefore, the better fit this representation is, the better our measure will be. In this sense, the practicality of ordinal scales is caused by them not only accommodating dimensions very well, but also interval and ratio dimensions. This aspect was very well addressed by Stevens (1946), where the upper scale levels must contain the previous levels. The possibility of performing an interval measurement indicates that psychologically, the elements of this continuum are processed additively and not positionally.

Thus, we searched for a more abstract (and possible "more cognitively") approach to the concept of color, and for that, we did not use color samples. A small sample was tested on a previous study and evidenced the applicability of this method for our purposes (Costa and Gaddi, 2016). In using physical stimuli, we were facing limiting factors to acquiring concept information in a more cognitive sense. Thus, in this paper we aimed to objectively measure the concept of the color for a higher level of abstraction by scaling their relative distance in the scale, assuming that its amount of salience is expressed on the order of color terms presented by the subject. There is no unique method that performs this analysis, so, by using a double methodology, frequency data at each position in rank could be transformed into mathematical units of distance and these are used for the construction of a one-dimensional interval scale of concept salience for colors. Additionally, we contributed to the scarcity of studies on color naming in children, allowing important insights into the development of this function.

2. Methods

2.1. Subjects

A hundred thirty-eight volunteers (mean age = 22.7 yrs; SD = 1.9; 62 females) with normal or corrected to normal visual acuity took part in the experiment. All participants were Portuguese native speakers. They had no visual complaints and normal color thresholds previously checked by the Trivector protocol of the Cambridge Color Test (M.F. Costa, Ventura, Perazzolo, Murakoshi and Silveira, 2006; Regan et al., 1994). Subjects had no formal training in color science, psychophysical experiments, or professions that deal with color directly such as architecture, plastic arts, or the like. They were opportunely recruited among the students and staff of the University of São Paulo in a proportion of 3:1. The students are from high and middle-high socioeconomic income and the staff members are almost from the middle socio-economic income and secondary educational level. The majority of them were born in São Paulo state.

An additional group of children was evaluated with ages ranging from 6 to 10 years (mean age = 8,9yrs; SD = 2.3), recruited among students of EMEF General Liberato Bittencourt and Maria Aparecida Broca Meirelles two primary public school of São Paulo city, also from the middle socio-economic income and parents with a secondary educational level. All children had normal or corrected to normal visual acuity and no ophthalmological complaints. The evaluation of the children took place on the premises of the school itself, in an isolated room intended only for our evaluation. The children participated in pairs, according to the convenience decided by the teachers.

The study was approved by the Ethics Committee for Human Research of Psychology Institute of the University of São Paulo. Informed consent was obtained for all participants and from the children's parents. The study followed the American Psychological Association research report and the tenets of the Declaration of Helsinki (1964) and all the posterior updates.

2.2. Procedure

The same procedure was done by children and adults. No visual stimulation was used and the task was extremely simple. The experiment was run in a quiet room, well illuminated in which all walls were painted off-white and the desks were covered by black paper. All pens were of black ink. The instructions were offered for small groups of subjects varying from two to five volunteers with the same number of subjects for the experimental session. During the instruction and experimental session, the researchers did not mention any term related to color. All subjects only were aware that was an experiment related to color at the instant of giving their answers. Each subject participated in only one session and the group composition occurred opportunely according to their availability to experiment.

The task performed by subjects was really simple and consisted in write a list of colors, in a column fashion, as fast as they were able in a completely white sheet of paper offered by the researchers. The participants were previously aware that the time to complete the experimental session was of 20 s. To control same-class color replications by using superlatives or compounding descriptions only names/words with one term were accepted. Previous studies found between 28 and 34.7% of color terms with more than one term color descriptor (Frank, 1990; Guest and van Laar, 2000). After all, the participants confirmed the understanding of the instructions they were aware that the name/words were about colors, and immediately after the experiment session began. As the session ends, the researchers inform the participants, asking them to stop, and no allowing more names/words then the paper sheets were collected. For children, the only adaptation consisted of that verbal information was collected instead of writing color terms, since part of the students were in the literacy process.

2.3. The psychophysical scaling

We used two methods developed by Thurstone in conjunction to construct an interval scale of concept saliency to colors. First, the data were analyzed based on the psychophysical Rank Order Scaling Method proposed by Louis L. Thurstone – for a detailed step-by-step procedure see (Thurstone, 1931b). The main outcome of that methodology is converting the ranking order in a correlational matrix aiming to construct an interval scale for one-dimensional continuous. This method is particularly appropriate when the dimension to be evaluated does not have a clear physical continuum available but we still want to build an interval scale from data that originally represented in rank order-type comparisons. We could consider this method as an ideal form of the classical constant method, in which we use not one but all the stimuli as the standards. From the data of absolute rank order, we should calculate the proportion of judgments "B is greater than A" for every possible pair of stimuli in the given series. These derived proportions are used instead of the proportion obtained by the constant method procedure.

If the participant ranks *n* color terms we tabulate them, in which each of the color terms could be compared with every other one in the series. In a more general form, if there were *n* stimuli in the color ranked it would require $n(n-1)$ judgments. This would give one judgment for each pair of colors possible. The data preparation and calculation follow some steps. The first step was to get the rank of color descriptors obtained for each subject. Since the participants were instructed to write the color terms in a column fashion the order from top-to-bottom lines we considered the first line as rank one, the second line as rank two, and so on. The second step consisted of calculating the proportion in which the color A was placed in the first rank position by all the participants (p_{a1}) and the proportion in which color B was placed in the first position (p_{b1}). Then, we calculate the sum of the probabilities for color A in all other rank positions, $p_{a2}+p_{a3}+p_{a4} \ldots. = p_{a>1}$. The same was done for other colors. Thus, the $p_{a1} * p_{b>1}$ is the probability of each participant, randomly chosen, place the color A in rank 1 and the color B in all other higher ranks. The same comparison was performed by color *A* for all other rank positions. The general product for all colors was written as the probability that color *A* will be perceived (P_e) in rank *K* and that color *B* will be perceived in a rank higher than *K*

$$p_{ak} * p_{b>k} = P_e \quad \text{(Eq. 2)}$$

Summing the total number of ranks for colors A and B we have the probability that color B will be perceived in a rank higher than that of color A

$$\sum (p_{ak} * p_{b>k}) = V \quad \text{(Eq. 3)}$$

Considering that two-color terms differ slightly in psychological value (*V*), so that B > A we would expect that $p_{a>b} > 0.50$.

Now, we could measure the probability that colors *B, C, D,* etc. that were perceived in a higher position than the rank of the color *A*. The general description to which each pair of all colors were placed in the *k* rank orders higher than the others can be formulated as

$$p_{b>a} = ©(p_{ak}*p_{b>k}) + ½\, ©(p_{ak}*p_{bk}) \qquad \text{(Eq. 4)}$$

Since in a rank order experiment we shall assume that participant number is as large as 10 or 20, the interval in a value represented by one rank order is relatively small. In such situations and especially when the discriminant error between terms is much larger than the interval represented by one rank order, the assumption is approximately correct, namely that probability that both A and B will be perceived in the same rank order interval and that B will be perceived higher than A. To correct that the ½ ©($p_{ak}*p_{bk}$) was added in the model expressing the proportion of participants who judged B higher than A in terms of the frequencies with which the two specimens are placed in the *n* rank orders. All those steps output tables are available in the Supplementary Material of this study. We could now build the third correlational matrix in which we have in columns the higher rank and in rows the lower rank positions considering the absolute values. Since we had the proportion of each color for each rank position, we could calculate the z-score value related to those proportions, which was converted in discriminative distances between colors. Finally, the color concept distances were calculated according to the Law Comparative Judgment (Thurstone, 1927a) using the correlational matrix resulted from the previous data analysis.

The main procedure to obtain an interval scale is the paired comparison, in which the participant is presented with two stimuli and asked to judge which is greater (or more intense) concerning some characteristic as redness, loudness, beauty, etc. Briefly, the pair comparison method could be summarized in three steps: first, presenting the participant each possible pair of stimuli asking "which has the highest (intense) appearance". This procedure is repeated until all the pairs were judged. It is interesting that, at least, two judgments being performed alternating the order (or position) for each pair randomly (Thurstone, 1927b). All of this step should be repeated for a single participant, which configures the Case I of the Law of Comparative Judgment or repeated for many participants, which configure the Case V. What is important is that the procedure is repeated until a reliable determination of the percentage of judgments could be made for each of the pairs. In the second step, we convert these percentages into "distances". One way to do it is to convert percentages in the area under the normal curve (z-score value), which expresses distances using standard deviation units (Gulliksen, 1946). The third step is to check if data satisfy the criterion for a linear scale by checking is the distances additively. If the third criterion was satisfied, the internal consistency is verified and the scale can be constructed.

The scale constructed by Thurstone's procedure is defined in terms of the frequencies of the comparative judgments on successive occasions about the same pair of stimuli (ideas, subjective values) for a particular psychological continuum. Thus, the psychological interval scale is spaced off that the frequencies of the successive comparisons assuming small fluctuation on judgments will generate small deviations from the equality, as well as huge asymmetry in judgments will produce large differences between the frequencies of one stimulus from the other. In other words, two almost similar stimuli will be closely located on the scale as well as two different stimuli will appear far apart from each other. Finally, the difference between the frequencies is the separation assigned to the two stimuli on the psychological scale on a distance scale as the Z distribution (Gulliksen, 1946).

The differences in scale values S between two stimuli *i* and *j* is

$$S_i - S_j = z_{ij}\sqrt[2]{\sigma_i^2 + \sigma_j^2 - 2\, r_{ij}\sigma_i\sigma_j} \qquad \text{(Eq. 5)}$$

where S_i and S_j represent the scale values of stimuli $_i$ and $_j$, $\int_i$ and $\int_j$ is the standard deviation of the respective discriminal dispersion, r_{ij} is the correlation between the two discriminal processes, and z_{ij} is the normal deviate (the Z-score) corresponding to the proportion of times stimulus *j* is judged greater along the psychological continuum than stimulus *i*.

Assuming that the evaluation of one stimulus along the continuum does not influence the evaluation of the other in the paired comparison ($r_{ij} = 0$) and the dispersions are equals for all stimuli ($\int_i = \int_j = 1$), the law is reduced to its simplest form (case V):

$$S_i - S_j = z_{ij}\sqrt[2]{2} \qquad \text{(Eq. 6)}$$

We used Case V that is the simplest version of the Law of Comparative Judgment. Since the points in the interval scale are based on equality of intervals or differences in the one-dimensional continuum. Such conditions allow the use of more informative parametric statistics as mean, standard deviation, and product-moment correlation. The color frequencies were transformed to proportions of appearance and those proportions in each rank position for each color were compared with all other colors. According to Thurstone's description, colors with proportion lower than 0.025 and higher than 0.975 should be excluded from the computation, due to the impossibility to assume the linearity near the extremes (Thurstone, 1928). Since the color fuchsia had a very low frequency, and all the proportions values were lower than 0.025, we excluded from the analysis considering a small z-score that will be related to it.

2.4. Internal consistency of the scale

The internal consistency of the scale was calculated considering the discrepancy of order judgment and was measured by using the scale values as a basis for sets of calculated proportions. These proportions were then, compared with the proportions obtained experimentally. Since we know the scale values measured, let's say for stimulus A by equation two, we were able to calculate the difference between stimulus A and all other stimuli and then converted each of them from z-score units back into the corresponding proportion. Those proportions were considered as the "calculated proportion", which was compared with the corresponding experimental proportion (P_e) obtained in the correlational matrix. The same procedure was used in an experimental study of nationality preferences (Thurstone, 1928). Statistical differences for percentages were calculated using V-corrected Chi-square, in which there is a sample-size correction applied to the Chi-square statistic. It is asymptotically equivalent to the Pearson Chi-square statistic but for small sample sizes, it is more conservative than the Pearson Chi-square (Overall et al., 1987). The comparison between Children and Adults results were assessed by the Mann-Whitney *U* test.

3. Results

The number of colors evoked by each adult subject varied from five to 11 terms (mean = 7.25; SD = 1.39; mode = 7), and a total of 16 different colors were evoked in the whole experiment. Children evoked 4 to 7 color terms (mean = 5.25; SD = 1.45; mode = 5). The frequencies of the colors are presented in Table 1.

Table 1
Color terms frequency for adults and children.

Color Term	Adult	Children
Red	79.7 (n = 110)	91.9 (n = 125)
Green	87.0 (n = 120)	91.9 (n = 125)
Blue	92.8 (n = 128)	95.6 (n = 130)
Yellow	87.1 (n = 121)	89.7 (n = 122)
Black	65.9 (n = 91)	88.9 (n = 121)
White	53.6 (n = 74)	80.1 (n = 109)
Purple	60.9 (n = 84)	73.5 (n = 100)
Brown	32.6 (n = 45)	51.4 (n = 70)
Pink	55.8 (n = 77)	77.9 (106)
Gray	27.5 (n = 38)	69.1 (n = 94)
Orange	48.6 (n = 67)	69.1 (n = 94)
Violet	6.5 (n = 9)	5.1 (n = 7)
Mangenta	9.4 (n = 13)	15.4 (n = 21)
Anil	6.5 (n = 9)	2.2 (n = 3)

The distances for color concepts calculated by equation II are presented in Fig. 1, in which the higher the value on the interval scaling the greater the amount of the concept's salience. Considering the interval scaling characteristic some computations could be performed. The scaling distances are shown in Fig. 1 reveal three clusters of terms between blue and yellow, pink and gray, magenta, and violet for both adults and children.

The ranges of the concept distances within each cluster are very similar for adults, 0.717SD for blue-yellow, 0.934SD for pink-gray, and 0.728SD for magenta-violet. The distances inter clusters also could be calculated, in which yellow-pink is .287SD is almost half-distance than gray-magenta is 0.511SD. Children's concept distances were 0.654SD for blue-yellow, 0.881SD for pink-gray, and 0.597SD for magenta-violet, which are quite similar to the adult values. However, the inter-clusters were 0.495SD for yellow-pink and 1.479SD for gray-magenta.

In order to compare the percentage of color terms used in our study we graphically compare them with the percentage reported for color terms of other two studies (Frank, 1990; Guest and van Laar, 2000). Statistical higher percentages were found by V-corrected Chi-square between our study and the Guest's study for the following colors: blue ($V^2 = 9.03$; $p = .003$), red ($V^2 = 11.69$; $p = .001$), green ($V^2 = 4.80$; $p = .028$), yellow ($V^2 = 50.97$; $p < .001$). Statistical lower percentages were for the colors: brown ($V^2 = 9.77$; $p = .002$), purple ($V^2 = 69.74$; $p < .001$), gray ($V^2 = 19.56$; $p < .001$), violet ($V^2 = 2.29$; $p = .130$). No difference was found for the colors: pink ($V^2 = 2.96$; $p = .085$) and orange ($V^2 = 3.29$; $p = .066$). The colors black and white were not presented by those authors (Fig. 2). Comparing our data with the Frank's study, statistical higher differences were found for the colors: blue ($V^2 =$

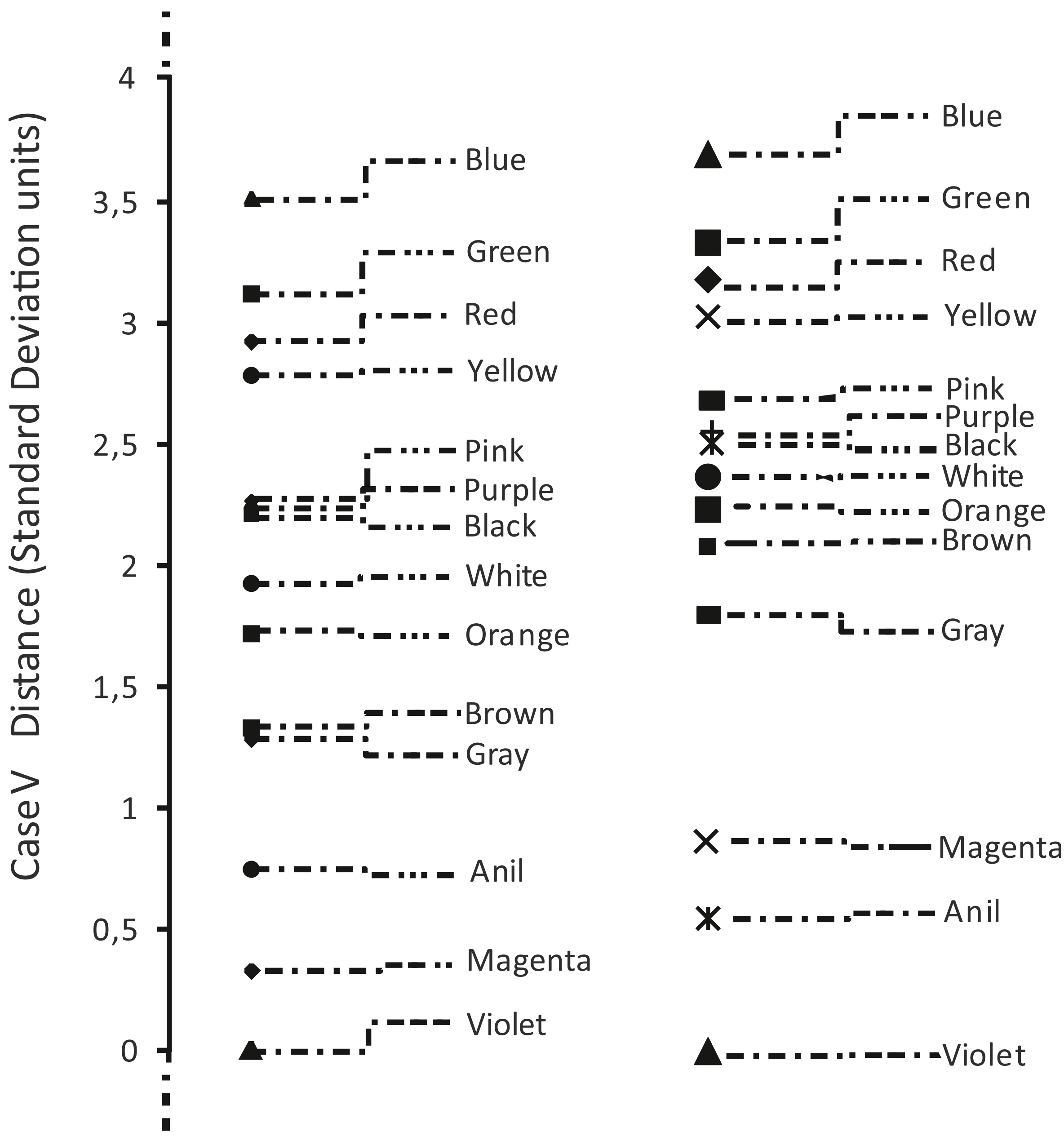

Fig. 1. Color concept scale obtained by Thurstone's Rank Order procedure for adults and children. The distances of the concept salience are expressed in the unities of the discriminal error (standard deviation units). Smaller the distances between the colors better the color concepts quantitatively determined. Based on these differences we can objectively say that in comparison with purple, magenta had 42 times less amount of salience than pink. Also, the order of colors are quite similar but the children scale shown larger distances than adults, more evident for the more complex colors concepts as magenta and anil.

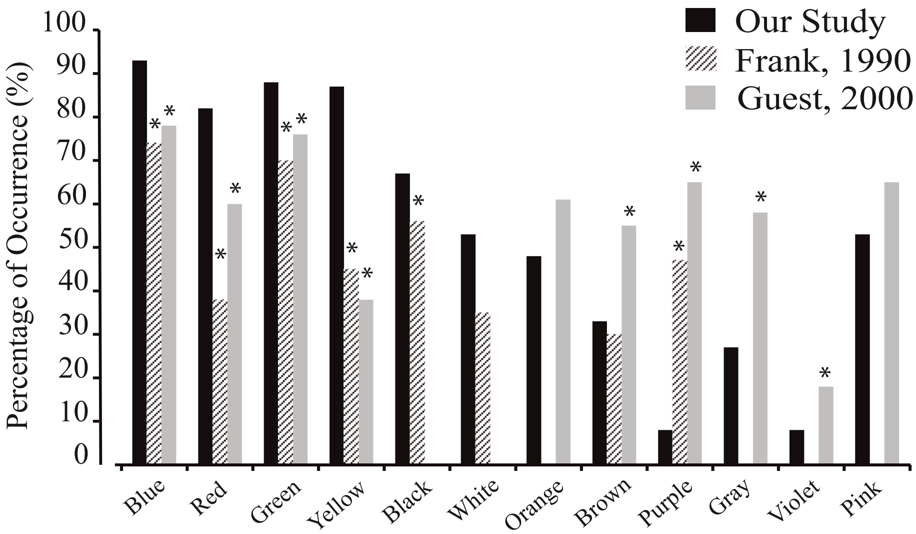

Fig. 2. The percentage of occurrence for the colors of our study was plotted together with the percentages reported in the studies of Frank and Guest. In comparison with Frank's paper, our study shows higher percentages with statistical significance for the colors: blue, red, green, and yellow. Lower percentages were found to brown, violet, gray, and purple. No difference between orange and pink. Compared with the Guest paper, higher percentages were found for the colors: blue, red, green, yellow, and black. The lower percentage was found in purple. No difference was found for brown.

13.04; $p = .002$), red ($V^2 = 40.13$; $p < .001$), green ($V^2 = 9.72$; $p = .002$), yellow ($V^2 = 39.11$; $p < .001$), black ($V^2 = 69.74$; $p < .001$). Statistical lower percentages were for the color purple ($V^2 = 37.95$; $p < .001$). No difference was found for the color brown ($V^2 = 0.21$; $p = .649$). The colors orange, gray, violet and pink were not presented by those authors (Fig. 2).

Additionally, we compare the rank position calculated by the Cognitive Salience Index and by Thurstone's method we used (Table 2).

Small positional changes in rank occurred for Yellow (CSI = 3; Case V = 4), Red (CSI = 4; Case V = 3), Pink (CSI = 8; Case V = 7), but larger changes for Black (CSI = 5; Case V = 8), White (CSI = 7; Case V = 5) colors.

Comparison between children and adults was verified for each color evoked for both groups. Significant statistical differences were found for White (Z-adj = 2.22; $p = .035$), Brown (Z-adj = 2.46; $p = .021$), Pink (Z-adj = 2.41; $p = .047$), Gray (Z-adj = 3.51; $p = .008$), Violet (Z-adj = 2.56; $p = .002$) and Orange (Z-adj = 2.47; $p = .020$) (Fig. 3).

The internal consistencies of the judgments were very high since variability between proportions calculated and obtained experimentally obtained was so low. For adults the variability of 0.001 was found for concepts blue, green, red, and yellow, 0.002 for pink, purple, black, and white, 0.004 for orange, 0.019 for brown, 0.021 for gray, 0.030 for magenta, and 0.055 for anil. For children the variability of 0.001 was found for concepts blue, green, red, and yellow, 0.002 for pink, purple, black, and white, 0.004 for orange, 0.019 for brown, 0.021 for gray, 0.030 magenta, and 0.055 for anil (Fig. 4).

Table 2
Comparison Salience Index and Case V results and their rankings.

Term	*Gloss*	CSI (*S*)	R	Case V	R
Azul	Blue	0,152	1	3385	1
Verde	Green	0,137	2	2999	2
Amarelo	Red	0,129	3	2808	4
Vermelho	Yellow	0,119	4	2668	3
Preto	Pink	0,084	5	2173	8
Roxo	Purple	0,065	6	2140	6
Branco	Black	0,052	7	2118	5
Rosa	White	0,052	8	1845	7
Laranja	Orange	0,040	9	1649	9
Marrom	Brown	0,018	10	1282	10
Cinza	Gray	0,013	11	1239	11
Magenta	Mangenta	0,002	12	0,728	12
Violeta	Anil	0,001	13	0,322	14
Anil	Violet	0,001	14	0,000	13

CSI (*S*) = Cognitive Salience Index; Case V = Thurstone's Case V scale value; R = rank.
Position based on CSI index value.

4. Discussion

The present work can be considered a step further on the study of the color concept since we measured concept distances on an interval scale, using Case V of the Law of Comparative Judgment. Based on this methodology we were able to perform high-level quantitative measurements of the cognitive salience of concepts for colors. Our analysis promotes an evolution on the systematic and quantitative study of the color concept measuring interval distances instead of classification, naming, or categorization analysis performed in most of the previous studies (Bachy et al., 2012; Bornstein, 1985; Kay and Regier, 2003) even for semi-interval methods, such as Cognitive Salience Index (Sutrop, 2001). All of those measurements are allocated on the categorical or ordinal measurement scale classifications and, consequently, have a lower level of information than those for metathetic stimuli (Stevens, 1958). Not only the concept rank position but the quantitative meaning of this position could be obtained by our method. Since we are using interval scale we can say that white were 1.83 times fewer salients than blue, while the anil was 10 times less salient, and white is 5.7 times more salient than the anil. This kind of analysis was not performed until now and which is the main contribution of our study.

This contribution is more evident by comparing the results obtained in adults and children. Similarly, based on the color blue we can say that white is 1.53 times less salient, while indigo is 36.7 times less salient than blue. White was 23.5 times more prominent than indigo. Thus, we do not have differences between the measured saliences for blue and white, showing that for the age we investigated, the concept of the blue and white color is developed when comparing the results of children with that of adults. However, there is a clear difference between blue and indigo and white and indigo, showing that the indigo color is still not very prominent for children. Our interpretation of these results is that secondary and tertiary colors are still in conceptual development.

Similar to adults the children's colors were relatively clustered in blue-green-red-yellow colors with higher salience, the intermediate group with pink-purple-black-white-gray, and finally the color with lower salience orange-brown-magenta-anil-violet. We could compare 2 information, the intra-cluster range and the between cluster range. In both, the smaller the range better was the concept of those colors, since small psychological concept distances suggest high conceptual knowledge about that color and with little distance change, the concept also changes to another color. As we compare the inter-cluster ranges, adults and children had similar results indicating that children with 9 years old

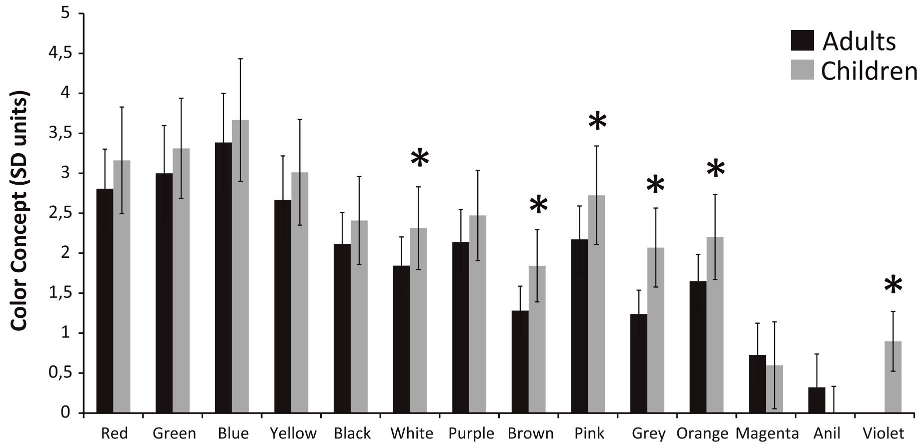

Fig. 3. The comparison between color terms evoked by children and adults. Significant differences are signaled, and occurred for complex color concepts as violet, orange and brown.

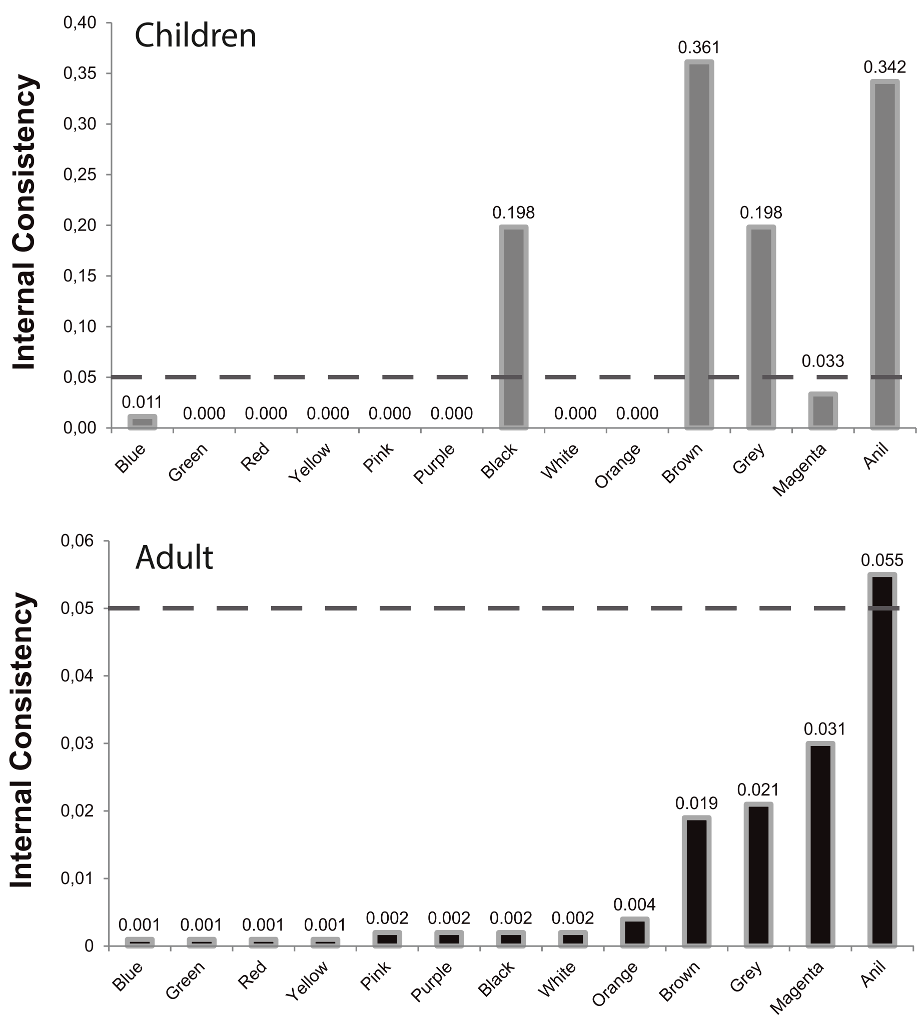

Fig. 4. Internal consistency was high for primary and secondary colors for adults and children. Dotted line defines the 5% of variability criteria. For tertiary colors almost all had also high consistency for adults. For children some of those tertiary colors had lower internal consistency (black, brown gray and anil).

had similar color concept development. However, the distance between the first and second clusters was two-fold longer for children and the second and third clusters almost three-fold longer. We could interpret these results as indicating that the color concepts between those clusters are still under development. For future studies, a comparison between different children's ages could elucidate how the concept for color development occurs.

Important evidence of our study is the possibility of using the methodology for the assessment of adults and children. In addition to the possibility of carrying out a quantitative assessment of conceptual distances, the possibility of applying them to children will allow future studies in which the development of this cognitive function can be explored more deeply. The evident characteristic of the use of this methodology is the association of a simple task but supported by robust psychophysics, allowing for measures of high internal consistency and possibility of application regardless of age.

One of the problems faced by psychophysicists is the long duration of the experimental sessions, due to the necessity of repeat many times the stimuli and also the necessity of sophisticated equipment, which made it difficult to evaluate huge populations or in their specific environment. Thus, Thurstone devised a method whereby simple absolute rank order was used as an experimental procedure with the similar advantages of the Method of Constant Stimuli and using the derivate proportion of the subjective separations between any pair of stimulus and then calculate the interval scale using the law of comparison judgment (Thurstone, 1931b). This method allowed a fast measurement with high psychometrical consistencies and reliability based on free-listing procedures.

The naming and color categorization studies carried out in various countries and cultures found concepts for color that were considered somewhat invariant and therefore elementary to our cognitive organization of this perceptual dimension (Lindsey and Brown, 2006), but differences in their relative importance could be exchanged between them as inferred by their position on the ranks, frequency of appearance or on the number of samples clustered to them. Since we can measure the amount of the concept salience along an interval scale, not only the differences between the ordering are collectible but the distances between them are also quantified. Also, the scale build by the Law of Comparative Judgment has the same dimension for the same procedure, which means that comparison between different populations, cultures, or ages could be performed since they were constructed on the same scale.

Our results are in line with the findings of previous works, since the primary, secondary and black, white, and gray colors were the colors more frequently evoked during the experiment as we showed in Fig. 1 (Derefeldt and Swartling, 1995; Ekici et al., 2006; Frank, 1990; Guest and van Laar, 2000). However, an important difference that we want to stress is that in our work was the ability to quantify and positioning the concepts of color along with a one-dimensional cognitive scale. This way of systematizing the concepts of color allows us not only to identify or categorize them but to measure the amount of salience that the concept presents for the group tested, based on the distance metric obtained by the psychophysical scaling.

The consistency of our results with those using naming and color categorization procedures lies in the fact that primary colors are found separately from secondary colors. Although, our clustering of data occurs not by any subjective quality of the color, as its saturation, as observed previously (Ekici et al., 2006) but solely by the amount of the concept salience since no sensorial stimulation was performed.

We present a methodological innovation on the usage of psychophysical scaling to derivate the interval scale of color concept. This innovative approach contributes with a quantitative measurement of more "pure" cognitive information regarding the concept of the color than the methodology used in previous studies, which were based on categorization or naming tasks that uses visual stimulation (Regier et al., 2007; Sivik and Taft, 1994; M.A. Webster and Kay, 2005). We believe that the method we used could promote new and future quantitative studies regarding the color concepts, concept formation, and the development of color concept topics without using visual stimulation and at the interval level of measurement. Our empirical experience allows us that our method applies to children of school age, with pertinent methodological modifications. (M. F. Costa and Gaddi, 2016). The methodology is simple but very substantial in the measurement of color concepts for use in different populations and different places. The significant internal reliability provides a powerful approach to that cognitive function.

The Method of Listing and their variants have been used as a tool for determining basic color terms and the data are analyzed by mean of the Cognitive Salience Index calculation, which is a function of frequency and priority (Smith and Borgatti, 1998; Sutrop, 2001; Thompson and Juan, 2006; Uuskula and Bimler, 2016). Considering the adjacency as an important variable, Uuskula & Bimler additionally calculated the adjacency index for every color term (Uuskula and Bimler, 2016) making it possible to build a separation matrix, which is similar to our calculated z-score distance matrix. However, the discrepancy between that study and ours is that, in our method, we used each term as the reference allowing the cross-correlation distance matrix aiming to map the color terms on a unidimensional interval scale.

The method of Rank Order is a simple task as a free-listing procedure and used as an experimental method produced quantitative interval data that could be used to construct a unidimensional scale allowing comparisons between different languages and cultures. The association of the Rank Order task with Thurstone's Case V based on the Law of Comparative Judgment further an objective calculation procedure for the concept's salience amount.

5. Conclusions

Our work contributes to the study of the concept of color compared to traditional studies of naming or categorizing colors since 1. We raise the metrics from categorical or ordinal levels to levels of interval measurements; 2. Quantitative color salience can be evaluated for adults and children using the same procedure allowing future studies in the development of color salience; 3. We decouple the measurements of color concepts from any sensory stimulation as an experimental reference, thus having a more cognitive/linguistic method than sensory-perceptual. Later studies investigating these diverse psychological representations can help in understanding how sensory and perceptual information is represented more conceptually. The simplicity and high precision of the procedure allow its applicability on large groups of different populations, in which comparisons are desired by the use of the same method.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank all the support received from the General Liberato Bittencourt and Maria Aparecida Broca Meirelles staff. MFC is a CNPq research fellow.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.metip.2021.100077.

Author contributions

Conceptualization: MFC; Data curation: MFC and FVP; Funding acquisition: MFC and FVP; Investigation: MFC, CMG, VMG, and FVP; Data analysis: MFC, FVP, and CMG; Writing - original draft: MFC and VMG; Writing - review & editing: all authors revised and approved the final version.

Funding

This work was supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo under Grant (FAPESP #13/04049-4); by the Projeto de Pesquisa Temático [FAPESP #14/26818-2]; Edital Ciências Humanas Conselho Nacional de Pesquisa, Tecnologia e Inovação [CNPq 404603/2018-0]; and the CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) to CMG (88882.376512/2019-01).

References

Arons, A.M.M., Krabbe, P.F.M., Scholzel-Dorenbos, C.J.M., van der Wilt, G.J., Rikkert, M. G.M.O., 2012. Thurstone scaling revealed systematic health-state valuation differences between patients with dementia and proxies. J. Clin. Epidemiol. 65 (8), 897–905. https://doi.org/10.1016/j.jclinepi.2012.01.018.
Athanasopoulos, P., Damjanovic, L., Krajciova, A., Sasaki, M., 2011. Representation of colour concepts in bilingual cognition: the case of Japanese blues. Biling. Lang. Cognit. 14 (1), 9–17. https://doi.org/10.1017/S1366728909990046.
Bachy, R., Dias, J., Alleysson, D., Bonnardel, V., 2012. Hue discrimination, unique hues and naming. J. Optical Soc. Am. A Optics Image Sci. Vis. 29 (2), A60–A68. https://doi.org/10.1364/JOSAA.29.000A60.
Bardovi-Harlig, K., 1987. Markedness and salience in second-language acquisition. Lang. Learn. 37 (3), 385–407.
Bornstein, M.H., 1985. On the development of color naming in young children: data and theory. Brain Lang. 26 (1), 72–93. https://doi.org/10.1016/0093-934x(85)90029-x.
Buchsbaum, G., Bloch, O., 2002. Color categories revealed by non-negative matrix factorization of Munsell color spectra. Vis. Res. 42 (5), 559–563. https://doi.org/10.1016/S0042-6989(01)00303-0.
Chapman, J.W., Tunmer, W.E., 1995. Development of young children's reading self-concepts: An examination of emerging subcomponents and their relationship with reading achievement. J. Educ. Psychol. 87 (1), 154–167. https://doi.org/10.1037/0022-0663.87.1.154.
Cieślicka, A., 2006. Literal salience in on-line processing of idiomatic expressions by second language learners. Sec. Lang. Res. 22 (2), 115–144. https://doi.org/10.1191/0267658306sr263oa.
Cole, B.L., Lian, K.Y., Sharpe, K., Lakkis, C., 2006. Categorical color naming of surface color codes by people with abnormal color vision. Optom. Vis. Sci. 83 (12), 879–886. https://doi.org/10.1097/01.opx.0000249974.22205.2a [doi];00006324-200612000-00008 [pii].
Costa, M.F., Gaddi, C.M., 2016. Color name distances scaled by Thurstone's ranking order psychophysical method. J. Vis. 16 (12), 199. https://doi.org/10.1167/16.12.824.
Costa, M.F., Nomura, T.N., Ribeiro, G.D.C., 2014. Psychophysical scaling of sexual attitude in Brazilian adults. Inpact 2014: Int. Psychol. Appl. Conf. Trend. 322–324.
Costa, M.F., Ventura, D.F., Perazzolo, F., Murakoshi, M., Silveira, L.C.D., 2006. Absence of binocular summation, eye dominance, and learning effects in color discrimination. Vis. Neurosci. 23 (3–4), 461–469 doi:10.10170S095252380623311X.
de Sousa, D.C.P., Soldati, G.T., Monteiro, J.M., Araujo, T.A.D., Albuquerque, U.P., 2016. Information retrieval during free listing is biased by memory: evidence from medicinal plants. PloS One 11 (11). https://doi.org/10.1371/journal.pone.0165838.
Derefeldt, G., Swartling, T., 1995. Color concept retrieval by free color naming - identification of up to 30 colors without training. Displays 16 (2), 69–77. https://doi.org/10.1016/0141-9382(95)91176-3.
Ekici, E.S., Yener, C., Camgoz, N., 2006. Colour naming. Opt Laser. Technol. 38 (4–6), 466–485. https://doi.org/10.1016/j.optlastec.2005.06.007.
Feher, E., Meyer, K.R., 1992. Childrens conceptions of color. J. Res. Sci. Teach. 29 (5), 505–520. https://doi.org/10.1002/tea.3660290506.
Frank, J., 1990. Gender differences in color naming: direct mail order advertisements. Am. Speech 65 (2), 114–126. https://doi.org/10.2307/455532.
Giora, R., 2003. On our mind: Salience, context, and figurative language. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195136166.001.0001.
Guest, S., van Laar, D., 2000. The structure of colour naming space. Vis. Res. 40, 723–734. https://doi.org/10.1016/S0042-6989(99)00221-7.
Gulliksen, H., 1946. Paired comparisons and the logic of measurement. Psychol. Rev. 53 (4), 199–213. https://doi.org/10.1037/H0061673.
Kay, P., Regier, T., 2003. Resolving the question of color naming universals. Proceed. Nat. Academy Sci. USA 100 (15), 9085–9089. https://doi.org/10.1073/pnas.1532837100 [doi];1532837100 [pii].
Krabbe, P.F.M., 2008. Thurstone scaling as a measurement method to quantify subjective health outcomes. Med. care 46 (4), 357–365. https://doi.org/10.1097/mlr.0b013e31815ceca9.
Li, Z., Cheng, K.E., Wang, Y., Hiltz, S.R., Turoff, M., 2001. Thurstone's law of comparative judgment for group support. In: Paper Presented at the Seventh Americas Conference on Information Systems (Massachusetts, USA).
Likert, R., Roslow, S., Murphy, G., 1993. A simple and reliable method of scoring the Thurstone attitude scales. Person. Psychol. 46 (3), 689–690. https://doi.org/10.1080/00224545.1934.9919450.
Lindsey, D.T., Brown, A.M., 2006. Universality of color names. Proceed. Nat. Academy Sci. USA 103 (44), 16608–16613, 0607708103 [pii];10.1073/pnas.0607708103 [doi].
Maydeu-Olivares, A., Böckenholt, U., 2008. Modeling subjective health outcomes: top 10 reasons to use Thurstone's method. Med. care 46 (4), 346–348. https://doi.org/10.1097/MLR.0b013e31816dd8d9.
Norman, C., Rieber, M., 1968. Facilitation of concept formation by children by the use of color cues. J. Exp. Psychol. 76 (3), 460–463. https://doi.org/10.1037/h0025579.
Overall, J.E., Rhoades, H.M., Starbuck, R.R., 1987. Small-sample tests for homogeneity of response probabilities in 2 X 2 contingency tables. Psychol. Bull. 102 (2), 307–314. https://doi.org/10.1037/0033-2909.102.2.307.
Pitchford, N.J., Mullen, K.T., 2001. Conceptualization of perceptual attributes: a special case for color? J. Exp. Child Psychol. 80 (3), 289–314. https://doi.org/10.1006/jecp.2001.2634.
Pitchford, N.J., Mullen, K.T., 2005. The role of perception, language, and preference in the developmental acquisition of basic color terms. J. Exp. Child Psychol. 90 (4), 275–302. https://doi.org/10.1016/j.jecp.2004.12.005.
Regan, B.C., Reffin, J.P., Mollon, J.D., 1994. Luminance noise and the rapid determination of discrimination ellipses in colour deficiency. Vis. Res. 24 (10), 1279–1299. https://doi.org/10.1016/0042-6989(94)90203-8.
Regier, T., Kay, P., Khetarpal, N., 2007. Color naming reflects optimal partitions of color space. Proceed. Nat. Academy Sci. USA 104 (4), 1436–1441. https://doi.org/10.1073/pnas.0610341104 ([doi]).
Rosch, E.H., 1973. Natural categories. Cognit. Psychol. 4 (3), 328–350. https://doi.org/10.1016/0010-0285(73)90017-0.
Sivik, L., Taft, C., 1994. Color naming - a mapping in the ncs of common color terms. Scand. J. Psychol. 35 (2), 144–164. https://doi.org/10.1111/j.1467-9450.1994.tb00939.x.
Smith, J.J., Borgatti, S.P., 1998. Salience counts—and so does accuracy: correcting and updating a measure for free-list-item salience. J. Ling. Anthropol. 7 (2), 208–209. https://doi.org/10.1525/jlin.1997.7.2.208.
Smith, J.J., Furbee, L., Maynard, K., Quick, S., Ross, L., 1995. Salience counts: a domain analysis of English color terms. J. Ling. Anthropol. 5 (2), 203–2016. https://doi.org/10.1525/jlin.1995.5.2.203.
Stevens, S.S., 1946. On the theory of scales of measurement. Science 103 (2684), 677–680. https://doi.org/10.1126/science.103.2684.677.
Stevens, S.S., 1958. Problems and methods of psychophysics. Psychol. Bull. 55 (4), 177–196. https://doi.org/10.1037/h0044251.
Sutrop, U., 2001. List task and a cognitive salience index. Field Methods 13 (3), 263–276. https://doi.org/10.1177/1525822X0101300303.
Taft, C., Sivik, L., 1997. Salient color terms in four languages. Scand. J. Psychol. 38 (1), 29–34. https://doi.org/10.1111/1467-9450.00005.
Thompson, E.C., Juan, Z., 2006. Comparative cultural salience: measures using free-list data. Field Methods 18 (4), 398–412. https://doi.org/10.1177/1525822X06293128.
Thurstone, L.L., 1927a. A law of comparative judgment. Psychol. Rev. 34 (4), 273–286. https://doi.org/10.1037/H0070288.
Thurstone, L.L., 1927b. The method of paired comparisons for social values. J. Abnorm. Soc. Psychol. 21, 384–400. https://doi.org/10.1037/h0065439.
Thurstone, L.L., 1928. An experimental study of nationality preferences. J. Gen. Physiol. 1, 405–426. https://doi.org/10.1080/00221309.1928.9918018.
Thurstone, L.L., 1929. The measurement of psychological value. In: Smith, T.V., Wright, W.K. (Eds.), Essays in Philosophy by Seventeen Doctors of Philosophy of the University of Chicago. Open Court, Chicago, pp. 157–174.
Thurstone, L.L., 1931a. The measurement of change in social attitude. J. Soc. Psychol. 2 (2), 230–235. https://doi.org/10.1080/00224545.1931.9918969.
Thurstone, L.L., 1931b. Rank order as a psychophysical method. J. Exp. Psychol. 14 https://doi.org/10.1037/h0070025.
Uuskula, M., Bimler, D., 2016. From listing data to semantic maps: cross-linguistic commonalities in cognitive representation of colour. Folklore Elect. J. Folklore 64, 57–90. https://doi.org/10.7592/Fejf2016.64.Colour.
Webster, M.A., Kay, P., 2005. Variations in color naming within and across populations. Behav. Brain Sci. 28 (4), 512–+.
Webster, M.A., Kay, P., 2012. Color categories and color appearance. Cognition 122 (3), 375–392. https://doi.org/10.1016/j.cognition.2011.11.008.
Zollinger, H., 1988. Categorical color-perception - influence of cultural-factors on the differentiation of primary and derived basic color terms in color naming by Japanese-children. Vis. Res. 28 (12), 1379–1382. https://doi.org/10.1016/0042-6989(88)90069-7.