This is a pre-print version of
Brychtova A., Coltekin, A., (2015) Discriminating Sequential and Qualitative Colour Schemes on Choropleth Maps, International Journal of Cartography, Vol xx:xx

Discriminating Sequential and Qualitative Colour Schemes on Choropleth Maps

Alžběta Brychtová*, Arzu Çöltekin**

* Department of Geoinformatics, Faculty of Science, Palacký University, Olomouc, Czech Republic
** GIVA, Department of Geography, Faculty of Science, University of Zurich, Switzerland

Abstract. We present an analysis of the influence of colour distance on the user experience with choropleth maps. We systematically evaluated 5 sequential and 5 qualitative colour schemes in a two-stage user experiment. At first, we conducted an online study to obtain performance metrics *accuracy* and *response time* on a large variety of heterogeneous population. Following this, in a controlled lab study with eye-tracking, we re-examined the findings from the online study for a subset of experimental stimuli and further assessed the user experience through an analysis of their visual behaviour. In this process, along with *accuracy* and *response time*, eye tracking metrics *fixation frequency, fixation duration* and *scanpath speed* as well as a *gaze transition analysis* were utilized. In both experiments, participants were asked to compare two areas with controlled colour distances between them, and decide whether these areas are of the same colour. Results suggest that we are able to discriminate narrower colour distances than commonly used in practice, however, not as narrow as suggested in previous literature.

Keywords: colour distance, usability, eye-tracking

1. Introduction

Colour is a very dominant visual stimulus, and therefore, one of the most important elements in map design. The use of colour in map design typically follows cartographic conventions based on centuries of practice and scholarship (e.g Slocum et al. 2008). Colour is also among the original seven *visual variables* proposed by Bertin and highly acknowledged in (and beyond) cartography since its publication (Bertin 2010, 1st edition 1967). When a visual variable is modified, it can have remarkable influence on user experience in tasks such as detecting change, or in reaction time on visual search tasks (Garlandini & Fabrikant 2009, Deeb et al. 2014). To control this influence, a *visual distance* between symbols can be introduced for each visual variable to make map use experience more productive and less error-prone (e.g., Bjorke 1996). For example, lack of proper visual distance in variables *colour hue* and *colour value* is a known contributor to legibility problems in map use tasks (Chesneau 2007, Stigmar 2010, Steinrücken & Plümer 2013). This visual distance between two shades of colour is termed *colour distance (*also referred to as *colour difference)*.

In this study, we systematically (re-)examine the influence of colour distance on the legibility of choropleth maps through a two-stage empirical user experiment. As common sense and previous studies suggest, we expect that increasing visual distance between colours will significantly improve the legibility of the map. We also expect to define a minimum threshold, below which human ability to detect different colours on thematic maps will be impaired. Our study contributes to the research efforts to better understand perceptually effective colour use in cartography based on empirical evidence, thus also contributes in user-centred design research.

2. Colour Distance on Maps

To quantify human ability to recognize the difference between colours, we employ the *colour distance* metric ΔE as introduced by International Commission on Illumination (CIE). Colorimetry scientists have been developing quantitative methods to accurately describe the colour distance as a metric and experimentally verifying whether two colours are distinguishable to the human eye. These efforts are also useful to define optimum colour differences for distinguishing map symbols in cartography. Additionally, to express colour quantitatively, specifications of various *colour spaces* have been progressively developed and refined over decades. Some of these colour spaces are optimized to correspond to the human perception. Such colour spaces are called *perceptually uniform* or *linear*, and their use ensures proportionality between numerically measured colour differences and the differences perceived by the human eye (CIE 2014). Presently, among the colour-distance models (or formulas), CIEDE2000 (ΔE_{00}) is regarded as the best coinciding with subjective visual perception. CIEDE2000 (ΔE_{00}) normalizes *brightness*, *hue*, and *saturation* of the visual perception to the same unit (Yang et al. 2012). When compared to other colour distance formulas, CIEDE2000 appears to be more reliable when applied to very small colour differences (<1) as well as very large colour differences (>10) (Carter & Huertas 2009).

A threshold at which it is *just* possible to differentiate between two visual elements is termed *Just Noticeable Difference* (JND). For colours, JND has not been precisely determined so far, i.e., various authors present slightly different values of JND for colour. For example, Yang et al. (2012)'s just noticeable difference is suggested at ΔE_{00}=0.5, while Linhares et al. (2008) propose ΔE_{00}=0.6.

Values close to JND are deemed inappropriate for cartographic purposes, because such small differences, especially when surrounded by other information, would make symbols undistinguishable. Therefore, cartographers typically apply higher values of colour. On the other hand, brighter colours with smaller (but still appreciable) colour distance better allow being overlain on other map information – e.g., labels, point or line features. Such features otherwise could disappear on a *too* dark background if the colour distance value is unreasonably high (Brychtová & Çöltekin 2014). Yet, in cartography, there have not been many attempts to empirically determine a threshold of the minimum *perceptually effective colour distance* or otherwise find the most effective values.

Steinrücken and Plümer (2013) have adopted the threshold of clearly distinguishable colours of 45 units of CIELUV[1] based on theory, however did not conduct a user study. In other studies, map colours that are created on demand seem to be mainly based on *colour-order systems* (e.g., Christophe 2008, Buard & Ruas 2009). The most significant contribution in the visualization domain regarding ordering colours appears to be the online software *ColorBrewer 2.0* and associated publications (Harrower & Brewer 2003). *ColorBrewer 2.0* provides specifications of various colour schemes of different shades and numbers of categories. The individual shades in *ColorBrewer 2.0* were picked from Munsell colour charts (Brewer 1989), which is perceptually informed but mathematically not as well-defined as, e.g., the CEIDE2000 or CIELUV.

To understand the current state of the art for commonly applied colour distances on maps and identify the minimum colour distance suggested by the most popular software tool; we examined 18 sequential colour schemes of six classes provided in *ColorBrewer 2.0* and calculated the theoretical colour distance (according to the CIEDE2000) between pairs of adjacent classes. Colour distances between them are not consistent among examined schemes (*Table 1*): the lowest colour distance among all schemes is

[1] CIELUV is abbreviation for CIE 1976 (L*, u*, v*) approximately uniform colour space adopted by the International Commission on Illumination. CIELUV values cannot be easily converted to CIEDE2000 units, because the transformation gives different results throughout the colour space.

ΔE_{00}=6.2; the highest ΔE_{00}=26.4. The most commonly applied colour distance ranges between ΔE_{00}=10 and ΔE_{00}=11.

ΔE_{00} range	(6-7]	(7-8]	(8-9]	(9-10]	(10-11]	(11-12]	(12-13]	(13-14]
Occurrences	4	5	9	5	10	9	9	7
ΔE_{00} range	(14-15]	(15-16]	(16-17]	(17-18]	(18-19]	(19-20]	(20-21]	(21-27]
Occurrences	3	3	8	2	4	4	3	5

Table 1. Occurrence of colour distance on 6-classed sequential colour schemes of ColorBrewer 2.0. Classes are most frequently differentiated by 10–11 units of ΔE_{00}.

3. Methods

A two-stage experiment was designed to examine the participants' ability to correctly distinguish two areas coloured by selected hues and values with a controlled colour distance between them. We hypothesize that increasing colour distance will have a consistent positive impact on distinguishability of visualized information. The first stage was conducted online, thus will be referred to as the *Web Survey* (WS). The second part of the experiment was carried out under controlled laboratory conditions with *eye-tracking*, thus will be referred to as ET for the remainder of the manuscript.

3.1. Participants

211 volunteers have participated in the online experiment (WS) (121 females, 90 males). Age groups were represented by 7% of 16–19 yr, 70% of 20–30 yr, 15% of 31–40 yr, 4% of 41–50 yr, 3% of 51–60 yr and 1% of more than 70 yr. The controlled lab experiment (ET) had 32 voluntary participants (19 females, 13 males). Age groups in the ET were 93% of 20–30 yr, 4% of 50–60 yr and 2% of 61-70. In both experiments, participants were asked to provide a self-evaluation of their expertise levels in cartography and GIS. In the WS, we obtained data from 54 male and 44 female experts, and 36 male and 77 female novices. Proportion of ET participants was rather similar: 9 male and 8 female experts and 4 male and 11 female novices.

3.2. Design and Procedure

Both studies were designed as within-subject factorial experiments with randomized stimuli. The WS was executed and analysed first to inform the ET experiment. The independent variable was the stimuli with controlled *colour distances* (see Section 3.3). For both experiments, our primary *dependent variables* were *accuracy* and *response time*. However, in ET, we also used eye movement metrics *fixation frequency, fixation duration, scanpath speed* and a *gaze transition analysis*.

There was a single task and it was designed to be purely perceptual: locate two areas on the map (marked with a dot), compare them and decide whether these areas were of the same colour. Participants could choose one of these three answers: (1) yes, marked areas are of the same colour; (2) no, marked areas are not of the same colour; (3) I don't know, I am not able to tell whether these colours are the same or not. No time limits were given.

The WS was implemented in the open source application *LimeSurvey* (The LimeSurvey project 2011) and participants were asked to keep the conditions 'as usual', i.e., under which they normally work with the computer to obtain some degree of ecological validity. However, they were also asked not to manipulate their screen settings and the room lighting during the survey to obtain some degree of control. The ET study was carried out under controlled conditions in the laboratory at Department of Geoinformatics, Faculty of Science, Palacký University in Olomouc, equipped with a low-frequency contactless eye-tracker *SMI RED 250* (SensoMotoric Instruments 2013) with a sampling frequency 120 Hz. Stimuli were projected on 23" LG Flatron monitor IPS231P. Screen was calibrated to sRGB colour space. Stimulus size was 1920x1080px. Experiment was prepared and presented in *SMI Experiment Center™*. Fixation detection was performed through the *SMI BeGaze™* using ID-T (dispersion threshold algorithm). Dispersion threshold was set to 50 px and a minimum length of 80 ms. Calculation of basic eye-tracking metrics has been performed in *OGAMA* (Voßkühler 2013) and statistical analysis of the data in the statistical software *R* (R Core Team 2013).

3.3. Materials

In order to obtain the same colour output on multiple digital displays, it is necessary to calibrate them and use viewing applications supporting ICC (International Color Consortium) profiles. In practice, this approach is not feasible for an online study. Therefore the calculation of colour distances refers to theoretical values: we first defined colours by sRGB primaries (standard sRGB IEC61966-2.1), we transformed them to CIE 1976 L*a*b* applying CIE standard illuminant D65, and finally we applied CIEDE2000 method to calculate the colour distances. The computation of colour distance and transformation between CIELAB and sRGB were done through a web calculator designed by Lindbloom (2012). Ten different colour schemes of six classes were examined: Five of them were *sequential* schemes (composed of six different shades of colour *value*); another five were *qualitative* schemes (six different shades of colour *hue*). Shades of both sequential and qualitative colour schemes were graded by uniform steps ΔE_{00}=2, 4, 6, 8 and 10. We selected *six shades of green* for sequential colour schemes and six colour hues with approximately constant lightness (*yellow, orange, red, violet, blue and green*) for qualitative colour schemes (see *Table 2* for specifications).

Sequential Colour Shemes

ΔE_{00}	category	L	a	b	R	G	B
10	A	94,8	-30	30	201	255	179
	B	79,45	-30	30	159	211	138
	C	66,1	-30	30	123	174	104
	D	54,67	-30	30	93	143	76
	E	44,67	-30	30	68	117	52
	F	33,12	-30	30	39	89	26
8	A	94,8	-30	30	201	255	179
	B	82,39	-30	30	167	219	146
	C	71,22	-30	30	136	188	117
	D	61,34	-30	30	110	161	92
	E	52,63	-30	30	88	138	71
	F	44,58	-30	30	67	117	52
6	A	54,71	-30	30	93	143	76
	B	61,34	-30	30	110	161	92
	C	68,63	-30	30	129	181	110
	D	76,63	-30	30	151	203	131
	E	85,4	-30	30	175	228	154
	F	94,8	-30	30	201	255	179
4	A	66,15	-30	30	123	174	104
	B	71,24	-30	30	136	188	117
	C	76,65	-30	30	151	203	131
	D	82,4	-30	30	167	219	146
	E	88,51	-30	30	184	237	162
	F	94,8	-30	30	201	255	179
2	A	79,47	-30	30	159	211	138
	B	82,39	-30	30	167	219	146
	C	85,4	-30	30	175	228	154
	D	88,5	-30	30	183	237	162
	E	91,7	-30	30	192	246	171
	F	94,8	-30	30	201	255	179

Qualitative Colour Shemes

ΔE_{00}	category	L	a	b	R	G	B
10	A	94,7	7	1	255	235	236
	B	94,8	3	11	255	238	217
	C	99,2	-5	14	255	255	224
	D	95	-8	2	226	245	236
	E	94	-5	-10	216	242	255
	F	94	3	-9	234	237	253
8	A	95,5	6	1	254	239	239
	B	95,9	2	9	255	241	225
	C	99,3	-4	11	255	255	230
	D	95	-5	1	231	244	236
	E	95	-4	-8	224	244	254
	F	95	2	-7	238	240	253
6	A	95	4	1	249	238	237
	B	96	2	6	252	242	229
	C	99,5	-2	7	255	255	238
	D	95	-4	1	234	243	237
	E	95	-3	-6	229	243	250
	F	95	2	-6	239	240	249
4	A	95	2	0	246	239	238
	B	96	1	4	249	243	234
	C	99,7	-1	4	255	255	245
	D	95	-2	1	237	242	238
	E	95	-2	-4	233	242	246
	F	95	1	-4	240	240	246
2	A	95	-1	0	238	242	238
	B	95	-1	2	241	241	235
	C	99	0	1	252	252	247
	D	97	1	0	248	246	244
	E	95	1	-2	241	240	242
	F	95	-1	-2	238	241	242

Table 2. Examined colour schemes' specification in CIELAB and RGB primaries.

On each stimulus, two areas were selected and marked with a dot (*Figure 1*). The colour distances between the selected areas were controlled to be same number of ΔE_{00}= 2, ΔE_{00}= 4, ΔE_{00}= 6, ΔE_{00}=8, ΔE_{00}= 10 or ΔE_{00}=0 (same colour). The spatial distribution of these areas were also controlled; to simulate various conditions that could occur in map reading, we distributed the two areas over the map: they could be next to each other, at mid-distance on the map, or at two extremes. Spatial distribution of all other colours was randomized to avoid (or distribute) *simultaneous contrast* effect as much as possible; even though it is important to note that this effect cannot be completely eliminated (Bláha & Štěrba 2014).

We used 106 stimuli in the WS and 41 in the ET experiment. The stimuli used in the ET are a subset of the ones from the WS.

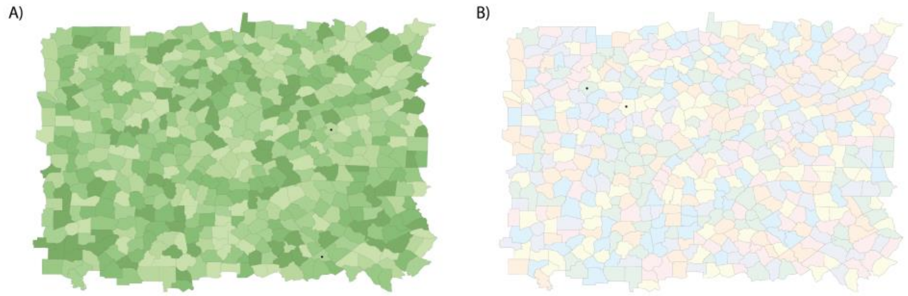

Figure 1. Example of experimental stimuli: sequential colour scheme with ΔE_{00}= 4 (A); qualitative colour scheme with ΔE_{00}= 8 (B).

4. Results

We first analysed the traditional performance metrics *accuracy* and *response time* both for the WS and the ET and compared the results from the two (one has a large number of participants, other has more control). Following the performance metrics, we analysed various eye-tracking metrics and gaze transition matrices. The hypotheses of the study were as follows:

(H1) Increasing colour distance will lead to a more accuracy

(H2) Increasing colour distance will result in shorter response time

(H3) Sequential and qualitative colour schemes will have no difference in accuracy or response time at the same colour distance level

(H4) Smaller colour distances will cause: longer average fixation durations (indicates difficulty in extracting information), higher frequency of fixations (indicates less efficient searching), longer scanpaths (indicates less efficient searching), and an increasing number of revisits between compared areas.

4.1. Accuracy

Both for the WS and the ET, the accuracy was summed up for groups of stimuli with the same colour distance, while sequential and qualitative colour schemes were analysed separately. We coded responses unsuccessful (i.e., inaccurate) when participants marked a difference between colours while there was not, and vice versa. “I don’t know” responses were also coded unsuccessful, because this suggested people could not tell if the two colours were identical or different, thus failed at the given task.

Analysing the coded responses based on the Pearson’s chi-squared test of independence revealed that accuracy was *dependent on colour distance in* *all examined conditions*: the WS sequential schemes (X^2= 453.27, df = 4, p<2.2*10^{-6}), the WS qualitative schemes (X^2= 254.19, df = 4, p<2.2*10^{-6}), the ET sequential schemes (X^2= 133.90, df = 4, p< 2.2*10^{-16}) and the ET qualitative schemes (X^2= 32.17, df = 4, p<1.62*10^{-6}). Additionally we found that accuracy of answers either from the WS (X^2= 15.47, df = 4, p=0.003) or from the ET (X^2= 58.57, df = 4, p=5.77*10^{-12}) are *dependent on type of colour scheme - sequential or qualitative*.

The results of the WS show a fairly consistent positive effect of increasing colour distance on the accuracy of answers (*Figure 2A*), as expected. Also, in the WS, both in sequential and qualitative colour schemes, the highest accuracy was observed at the largest ΔE_{00} = 10 (98% and 96% respectively) and the most troubling colour distance was the smallest ΔE_{00} = 2 (78% and 80% respectively). Interestingly, the ET experiment does not have a consistent positive effect of increasing colour distance on accuracy of answers as WS did. Results show that the most troublesome colour distance for the sequential scheme is as expected the ΔE_{00}=2 (58%), while for the qualitative scheme it is ΔE_{00} = 4 (82%) (*Figure 2B*). To understand whether this finding is because of an experimental artefact, we studied the input stimuli closely and observed that of the five stimuli at distance ΔE_{00} = 4 in the ET for qualitative scheme, three of them had 100% accuracy while one had 65% and another one 50%. The 50% success was with a stimulus that likely suffers from simultaneous contract effect, thus explains the odd finding in this particular condition. This particular stimulus also has a low success rate in the WS, however, because we have a larger sample size, it does not express itself as strongly in the average as in the ET experiment.

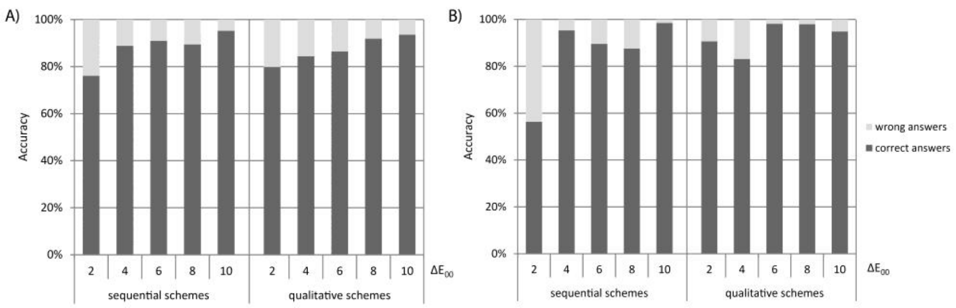

Figure 2. The accuracy rates for the WS (A) and the ET (B).

4.2. Response time

Overall, in the WS, questions were answered remarkably slower than in the ET. For the WS, the response times (for a single task) varied from 2.37 to 920.88 s while for the ET this is from 0.18 to 88.9 s. However, this is mainly due to the differences in the experiment procedure: For WS, the response time includes task solving *and* marking the response (they were displayed on the same page). For the ET, the answer form was displayed separately from the stimulus and only the stimulus viewing time was recorded (not the time spent marking the response). Additionally, extremely high values at the WS cannot be attributed to the performance issues. Since the online study is not controlled, it is likely that some participants took a break, paid attention to another task or were otherwise interrupted. Eye-tracking experiment was conducted in a controlled environment, and therefore no interruptions occurred. However, these differences between the WS and ET are not important; because only the relative differences between the tested conditions (colour distances and colour schemes) are meaningful. Nonetheless, to avoid possible bias from the interruptions in the WS; tasks that took longer than 90 seconds (based on the maximum response time observed in the ET), i.e., 65 outliers (not participants but individual tasks), were removed from the WS.

Based on the remaining recordings; for the correct answers, median response times for the WS was *Mdn=6.79 s* (IQR=9.34-5.33) and for the incorrect answers, it was *Mdn=8.18 s* (IQR=12.15-5.95). For the ET – i.e., for correct answers median was *Mdn= 2.64 s* (IQR=4.86-1.70) and incorrect answers it was *Mdn=3.49* s (IQR=6.20-1.93). In both experiments, participants took more time to mark the wrong answer than to mark the correct one. Median difference is *1.39 s* and *0.85 s* for WS and ET respectively.

Response times for the WS were further analysed with Kruskal-Wallis test based on correct answers only. Kruskal-Wallis tests examines with the null hypothesis that all populations have identical distribution functions against the alternative hypothesis that at least two of the samples differ only with respect to location of median. An analysis for the main effect yields statistically significant results: WS sequential (H=184.27, df=4, $p< 2.20*10^{-16}$), WS qualitative (H=52.29, df=4, $p=1.03*10^{-10}$), ET sequential (H=16.44, df=4, p=0.002) and ET qualitative (H=35.44, df=4, $p=3.77*10^{-7}$). Post-hoc Kruskal-Wallis test revealed that participants performed significantly worse when working with $\Delta E_{00}=2$. For the WS, both sequential and qualitative colour schemes have median time significantly higher with $\Delta E_{00}=2$ than any other ΔE_{00}. The difference is over 1 second for the sequential colour scheme ($\Delta E_{00}=2$ median is *Mdn=7.97 s*, while other colour distances have their medians lower than 6.96 s); and about 0.5 for the qualitative colour and significant as well (*Table 3*).

	Sequential color schemes		Qualitative color schemes	
ΔE_{00}	*Mdn*	*IQR*	*Mdn*	*IQR*
2	7.97	11.30–5.93	6.90	9.68–5.43
4	6.80	9.47–5.32	6.67	9.13–5.29
6	6.80	9.17–5.38	6.48	8.79–5.12
8	6.61	8.98–5.18	6.47	8.61–5.13
10	6.96	9.36–5.56	6.58	8.80–5.22

Table 3. Median values (Mdn) and Interquartile distribution (IQR) for the WS for colour distances ΔE_{00} =2–10 and colour scheme types *sequential* and *qualitative*. Vertical lines depict statistically significant differences between pairs.

Response times for ΔE_{00}=8 and 10 on sequential colour schemes also shows statistical significance, while no other colour distances lead to faster or slower response times. Thus increasing colour distance in a step-wise fashion did not cause a consistent decrease in response times. However, our WS results clearly suggest that colour distance ΔE_{00}=2 severely decreases participant performance (both in accuracy and in response time), and should be avoided in choropleth maps (both sequential and qualitative).

Overall, the results obtained from the ET support the WS results. For the qualitative schemes, we see even more statistically significant results, i.e., between ΔE_{00}=4 and 6 and ΔE_{00}=4 and 10. In this case difference between ΔE_{00}=2 and 4 is not statistically significant, however, these two distances seem to be the more time-consuming than others (see the *Table 4*).

	Sequential color schemes		Qualitative color schemes	
ΔE_{00}	*Mdn*	*IQR*	*Mdn*	*IQR*
2	4.10	8.54–2.50	3.02	5.53–1.98
4	2.67	4.66–1.85	2.69	4.92–1.59
6	2.81	5.96–1.74	1.98	3.51–1.38
8	2.94	4.68–1.97	2.03	3.65–1.57
10	2.97	5.49–2.05	1.94	3.32–1.45

Table 4. Median values (Mdn) and Interquartile distribution (IQR) for the ET for colour distances ΔE_{00} =2–10 and colour scheme types *sequential* and *qualitative*. Vertical lines depict significant differences between pairs.

We also hypothesized that the same level of colour distance would yield similar performances between the sequential and qualitative colour schemes. A Mann-Whitney test for the response times in the WS yielded statistically significant results between the sequential and qualitative schemes for ΔE_{00}=2 (U=1375163, $p=2.20*10^{-15}$), ΔE_{00}=6 (U=1958850, $p=0.01*10^{-2}$) and ΔE_{00}=10 (U=2176302, $p=2.98*10^{-7}$). In all cases, median time of sequential colour schemes is longer. The major difference is, when comparing ΔE_{00}=2: sequential scheme has 1.07 s higher median and also the interquartile range (IQR) is much wider than in qualitative schemes. The ET experiment fully confirms these findings; differences of medians and IQRs are even higher than observed for the WS and all comparisons except ΔE_{00}=4, are statistically significant.

4.3. Eye-tracking metrics

To better understand the reasons behind the findings based on the traditional performance metrics, we first analysed basic eye movement metrics: fixations frequency (number of fixations per second), average fixation duration (second) and scanpath speed (pixels per second). These metrics may indicate cognitive load or participant strategies during task execution (Holmqvist et al. 2011). Examples of eye movement studies that have confirmed the link between these metrics and cognitive load for map use also exist (e.g., Coltekin et al. 2009, Popelka & Brychtová 2013). Furthermore, we analysed gaze transition matrices, which report movements between defined areas of interest (gaze shifting back and forth between two areas), thus possibly indicating a struggle with the task (Holmqvist et al. 2011).

Fixation frequencies were not influenced by colour distance. A Kruskal-Wallis test suggested that the overall medians (of all examined colour distances) do not differ; neither in sequential (H=8.09, df=4, p=0.08), nor in qualitative colour schemes (H=6.55, df=4, p=0.16). A Mann-Whitney test further indicated that there is no significant difference between sequential and qualitative schemes at same colour distance levels. Overall median fixation frequency (without distinguishing colour distance or scheme type) is *Mdn=4.44 fixations/s* (IQR=4.90-3.89).Variations were observed for average fixation durations in sequential colour schemes (H=17.03, df=4, p=0.001. However, performing pairwise comparison, we found significant results only between ΔE_{00}=2 and 4 and ΔE_{00}=2 and 8. Median average fixation duration for ΔE_{00}=2 (*Mdn=191.30 ms*) is slightly higher than for ΔE_{00}=4 (*Mdn=167.45 ms*) and ΔE_{00}=8 (*Mdn=175.00 ms*), which may indicate more difficulties in extracting information from the map (Holmqvist et al. 2011). No variations were observed between average fixation durations on qualitative colour schemes (H=8.02, df=4, p=0.09).

Scanpath speed (pixels per second) analysis revealed many significant differences among sequential (H=97.46, df=4, $p<2.20*10^{-16}$) as well as qualitative schemes (H=47.63, df=4, $p=1.12*10^{-9}$). All differences found by pairwise comparison are visualized in *Figure 3* with linkers.

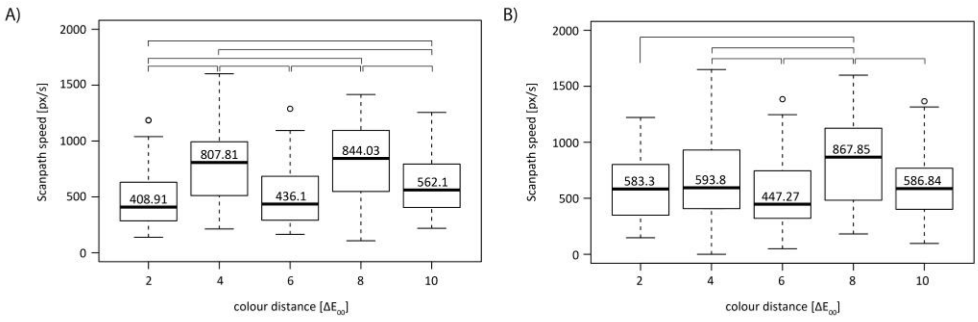

Figure 3. Scanpath speed [px/s]: sequential (A) and qualitative (B) colour schemes.

Highest median scanpath speed (*Mdn=867.85 px/s*) occurs on the qualitative colour scheme at ΔE_{00}=8. Similar results occur on sequential colour schemes, where the highest median values are for ΔE_{00}=8 and ΔE_{00}=4 (*Mdn=844.03 px/s* and *Mdn=807.81 px/s* respectively).

Analysing the gaze transition matrices supported our hypothesis, that lower colour distance will raise the number of back and forth gaze shifts between compared areas (*Figure 4*).

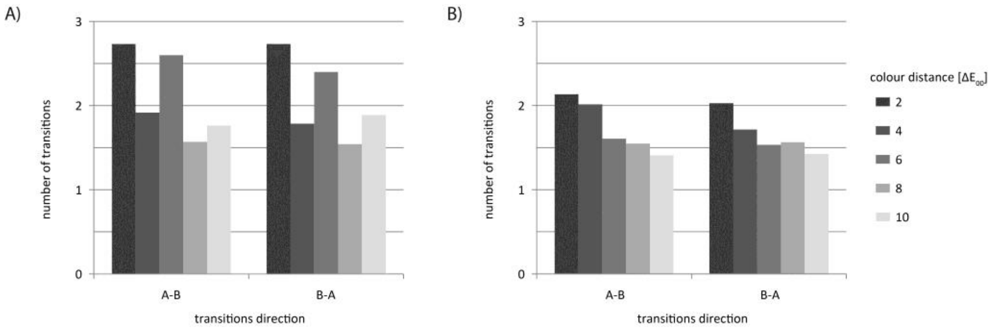

Figure 4. Average number of transitions per participant between area A and B: sequential (A) and qualitative (B) colour schemes.

Number of transitions between areas A and B (corresponding to colours to be compared) decrease in both directions (A-B and B-A) as the colour distance between classes of qualitative schemes grows. These results suggest that qualitative schemes may be easier to interpret than sequential schemes on the same level of colour distance: i.e., the average number of transitions is significantly higher in the case of ΔE_{00}=2 and 10 on sequential schemes.

5. Discussion and conclusions

We have systematically analysed the influence of colour distance on map readability through a two-stage user experiment. First experiment was conducted online to capture data from a wider population and then we performed eye-tracking study in a controlled laboratory to verify and better understand our findings. Participants were asked to locate two areas marked on a digital static map and decide whether these are of the same colour or not. We experimented with two types of a colour schemes (sequential and qualitative) and 5 levels of colour distance between adjacent classes (ΔE_{00}=2, 4, 6, 8 and 10).

We measured accuracy, response time and selected eye tracking metrics (fixation frequency, fixation duration, scanpath speed and gaze transitions).

We found that colour distance and type of colour scheme influence the accuracy of responses. The results of the web survey shows a positive effect of increasing colour distance on the map-users' ability to correctly compare visualized spatial information. For both sequential and qualitative colour schemes the highest accuracy was caused by largest colour distance ΔE_{00} = 10 (over 95% accuracy); the most problematic was ΔE_{00} = 2 (less than 80% accuracy). Overall, the eye-tracking experiment confirmed obtained results, however with some differences – which can be introduced by various differences between the two studies, i.e.; we used a subset of the stimuli (106 in the WS while 41 in the ET). Additionally the WS allowed marking the questions on the same display, while in the ET experiment participants had to remember what they have looked (even if it was right after they studied the visual stimulus) and mark the answer on the next screen. This may have caused some irregularity in participants' success rates. Last but not least, we had a remarkable difference in the number of participant between the two experiments (211 and 32 in the WS and in the ET, respectively). Another possible factor could be the fact that in the WS 211 participants have used unique displays with different colour configurations, while in the ET this was constant. While keeping the conditions constant is good for studying possible causes, in this case it may have meant that if there was a simultaneous contract effect, we also kept that in the stimulus more consistently (while e.g., this visual illusion may be weaker in some other display configurations).

Regardless of the minor differences between the WS and the ET, our results clearly demonstrated that the lowest examined colour distance ΔE_{00} = 2 led to more errors and slowed people down despite being marked as "noticeable" by Young et al. (2012). We can also safely say that ΔE_{00} = 10 is a safe colour distance, but the rest of the tested distances are essentially not recommendable. Even though the error rates are fairly low, the task was an extremely simple perceptual task, thus if the colours were distinguishable for all, it should have yielded near-zero error rates.

In this particular study, basic eye-tracking metrics (fixations frequency, fixation duration, scanpath speed) did not offer additional explanations. Perhaps this could be attributed to the simplicity of the search task (the two areas were marked each with a dot), that does not result in complex gaze trajectories. Number of gaze transitions between two areas was in average the highest with $\Delta E_{00} = 2$ (more than 2.5 transitions for sequential and more than 2 for qualitative schemes), validating the other findings.

Overall response time and gaze transition analysis showed that distinguishing between two areas is more difficult on sequential maps, than qualitative. This may be caused by the fact that people can better distinguish colours that are able to name (Brewer 1996). To remember shades of the same colour can be difficult, even impossible (Albers 2013), because the memory works only with visual perception without assigning specific semantic value (name of the colour), which would activate the verbal working memory to further support the visual processing.

References

Albers J (2013) Interaction of Color (p. 208). Yale University Press.

Bertin J (2010) Semiology of Graphics: Diagrams, Networks, Maps. Components (1st ed., p. 456). ESRI Press.

Bjorke JT (1996) Framework for Entropy-based Map Evaluation. Cartography and Geographic Information Systems, 23(2), 78–95.

Bláha JD & Štěrba Z (2014). Colour Contrast in Cartographic Works Using the Principles of Johannes Itten. The Cartographic Journal, 51(3), 203–213.

Brewer CA (1996) Prediction of Simultaneous Contrast between Map Colors with Hunt's Model of Color Appearance. Color Research and Application, 21(3), 221–235.

Brewer CA (1989) The development of process-printed Munsell charts for selecting map colors. The American Cartographer, 16(4), 269–278.

Brychtová A & Çöltekin A (in press) An empirical user study for measuring the influence of colour distance and font size in map reading using eye tracking. The Cartographic Journal, 51(4).

Buard E & Ruas A (2009) Processes for Improving the Colours of Topographic Maps in the Context of Map-on-demand. In Proceedings of the 24th ICC (p. 11). Santiago, Chile.

Carter R & Huertas R (2009) Ultra-large color difference and small subtense. Color Research & Application, 35(1), 4–17.

Chesneau E (2007) Improvement of Colour Contrasts in Maps : Application to Risk Maps. Proceedings of 10th AGILE International Conference on Geographic Information Science 2007, 1–14.

Christophe, S. (2008). Legend Design on the Web: Creating Accurate Styles. International Journal of Spatial Data Infrastructures Research, 3, 38–57.

CIE (2012) Termlist of International Commission on Illumination. Retrieved from http://eilv.cie.co.at/

Çöltekin A et al. (2009) Evaluating the Effectiveness of Interactive Map Interface Designs: A Case Study Integrating Usability Metrics with Eye-movement Analysis. International Journal of Geographical Information Science, 36(1), pp. 5–17.

Deeb R et al. (2014). Background and foreground interaction: Influence of complementary colors on the search task. Color Research & Application.

Garlandini S & Fabrikant SI (2009) Evaluating the Effectiveness and Efficiency of Visual Variables for Geographic Information Visualization. In S. K. Hornsby (Ed.), COSIT 2009, 195–211.

Holmqvist K et al. (2011) Eye Tracking: A comprehensive guide to methods and measures (1. ed., p. 560). Oxford University Press.

Harrower M & Brewer CA (2003) ColorBrewer.org: An Online Tool for Selecting Colour Schemes for Maps. The Cartographic Journal, 40(1), 27–37.

Lindbloom BJ (2012) Useful Color Calculators and Spreadsheets. Retrieved February 12, 2013, from http://www.brucelindbloom.com/

Linhares JMM, Pinto PD & Nascimento SMC (2008) The number of discernible colors in natural scenes. J. of the Optical Society of America, 25(12), 2918–24.

Popelka S & Brychtová A (2013) Eye-tracking Study on Different Perception of 2D and 3D Terrain Visualisation. The Cartographic Journal, 50(3), 240–246.

R Core Team (2013) R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/

SensoMotoric Instruments (2013) SMI Experiment Suite 360TM. Retrieved from http://www.smivision.com/en/gaze-and-eye-tracking-systems/products/ experiment-suite-360-software-package.html

Slocum TA et al. (2008) Thematic Cartography & Geographic Visualization, 3rd edn. Prentice Hall, Upper Saddle River

Steinrücken J & Plümer L (2013) Identification of Optimal Colours for Maps from the Web. The Cartographic Journal, 50(1), 19–32.

Stigmar H (2010) Making 21st Century Maps Legible - Methods for Measuring and Improving the Legibility and Usability of Real-Time Maps. Lund University.

The LimeSurvey project (2011) LimeSurvey - the free and open source survey software tool. Carsten Schmitz. Retrieved from https://www.limesurvey.org/

Voßkühler A (2013). OGAMA (OpenGazeAndMouseAnalyzer). Freie Universität Berlin. Retrieved from http://www.ogama.net/

X-Rite (2012) How Color Notation Works. Munsell Color. Retrieved from http://munsell.com/about-munsell-color/how-color-notation-works/

Yang Y, Ming J & Yu N (2012) Color Image Quality Assessment Based on CIEDE2000. Advances in Multimedia, 2012, 1–6.