PulseLight

Anna Sturm, Axel Bauer, Thomas Förg

Supervisor: Martin Kocur

Background & Motivation

We wanted to find out which display option on smartwatches best supports users during their sports activities because usually, the default display mode on an Apple Watch is text-only. Our thought is that especially when doing heavy workout - having tunnel vision - the perception of the most vital data, the heart rate, is not sufficient enough. Our approach is to turn the smart watch into a solely body-interface-device which gives the user more control over their workout and faster feedback without having a distraction when focusing on reading the values.

Related Work

Performance feedback in sports:

Gottschalk, M. (2013). Reinforcing feedback to improve performance in jumps is superior to external and internal focus of attention.

> *Figoni, S.F. & Morris, A.F. (1984). Effects of Knowledge of Results on Reciprocal, lsokinefk Strength and Fatigue. J Orthop Sports Phys Ther., 6(3), 190-197.*
>
> *Studie: Hopper, D. M., Axel Berg, M. A., Andersen, H. & Madan, R. (2003). The influence of visual feedback on power during leg press on elite women field hockey play-ers. Physical Therapy in Sport, 4(4), 182–186. doi:10.1016/S1466-853X(03)00068-3*

Color theory: Bellantoni, P. (2005). *If it's purple someone's gonna die, The Power of Color in Visual Storytelling* (e-book). Focal Press.

Research Question

Is there a difference regarding understanding and time to understand when it comes to the reception of BPM data on a smartwatch during workout between a display through text and/or color?

Study Design incl. Participants

- quantitative study - "between-subject" design - 3 display modes (IV)

Subject	1	2	3
#1	color	colored number	white number
#2	color	white number	colored number
#3	white number	color	colored number
#4	white number	colored number	color
#5	colored number	white number	color
#6	colored number	color	white number

https://hci-studies.org/balanced-latin-square/

- Since 6 participants are not enough, the participants were doubled the latin-square = 12 participants

Measures - Dependant Variables

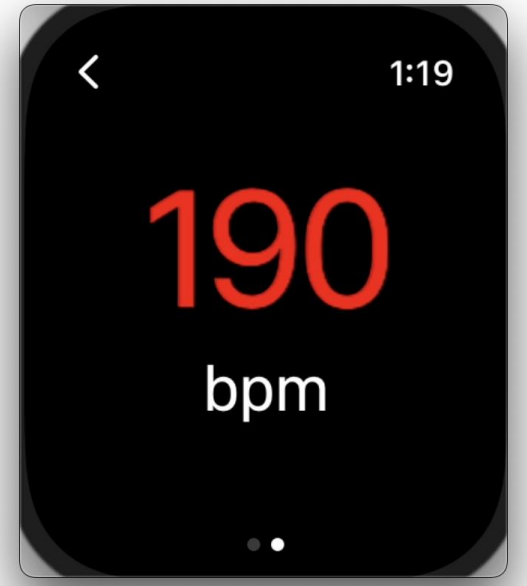
- *Error rate (of answers - possible answers: high, middle, low)*
- *Time until answer is given*
- *BPM (heartbeats per minute) of the participant - validating the BORG scale*

- BORG - Rating of perceived exertion
- SUS (System Usability Score) - Rating the usability of the display mode
- Demographics - In order to distinguish/group the participants

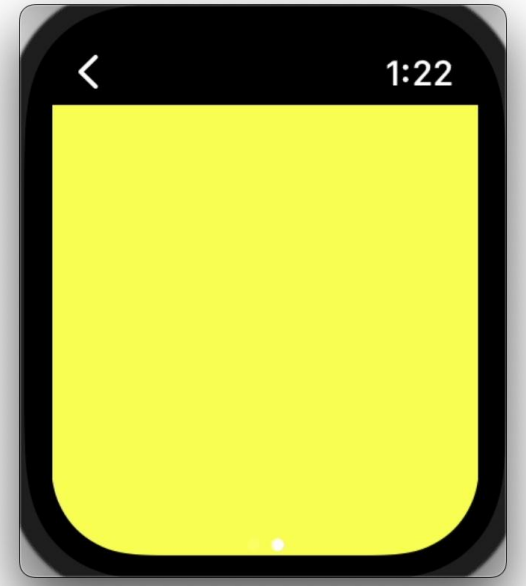
Apparatus - Independent Variables

Overview

Metrics View



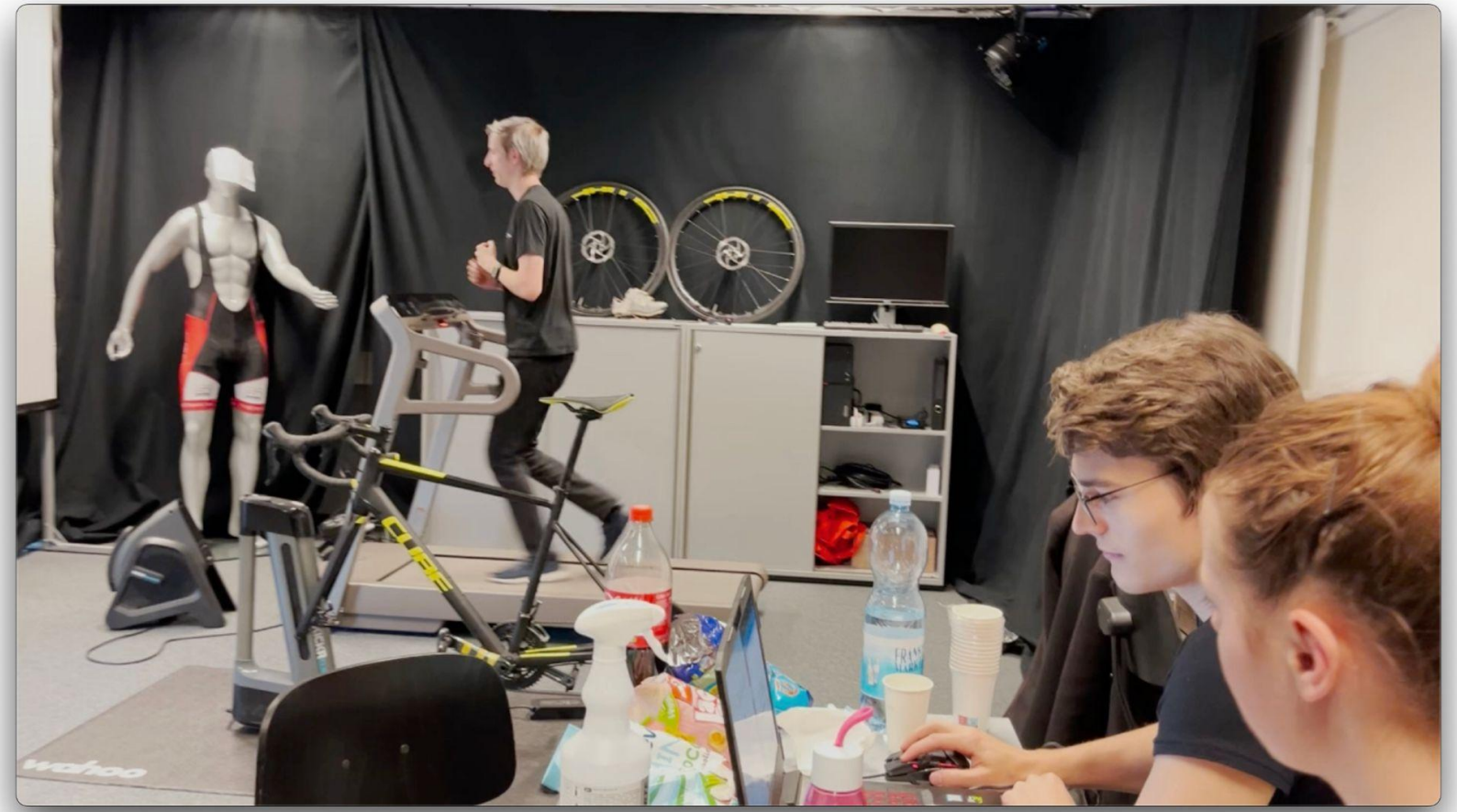
Colored Text View

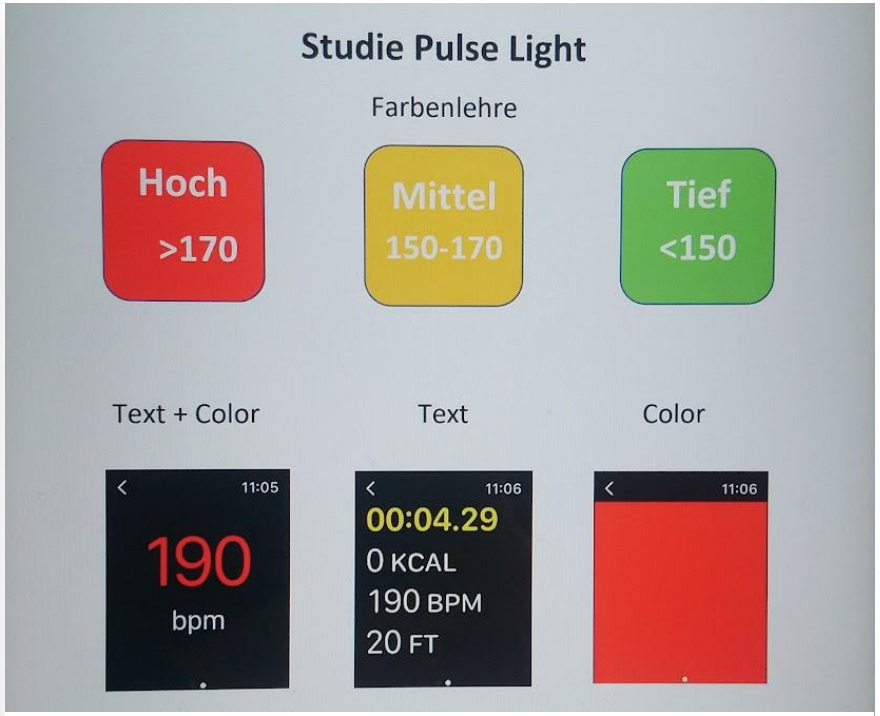


Color View

3 display modes / IVs

Apparatus - Study

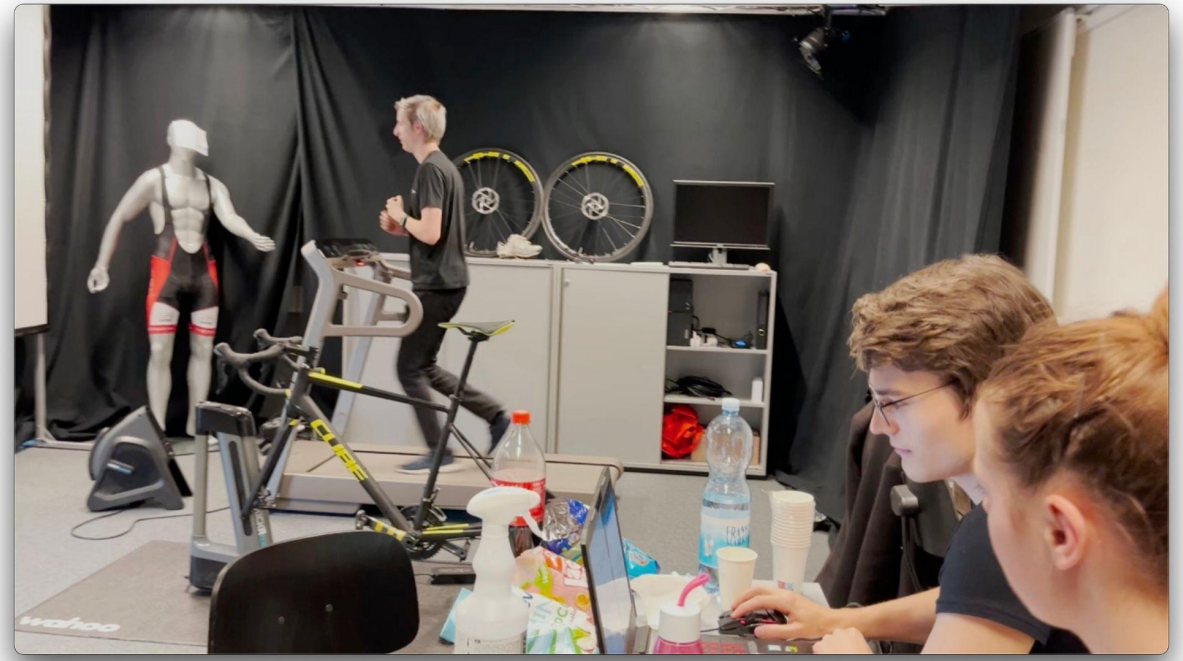




Tablet which displays the value ranges (color and BPM) and below screenshots of the watchOS app.

Apparatus - Study

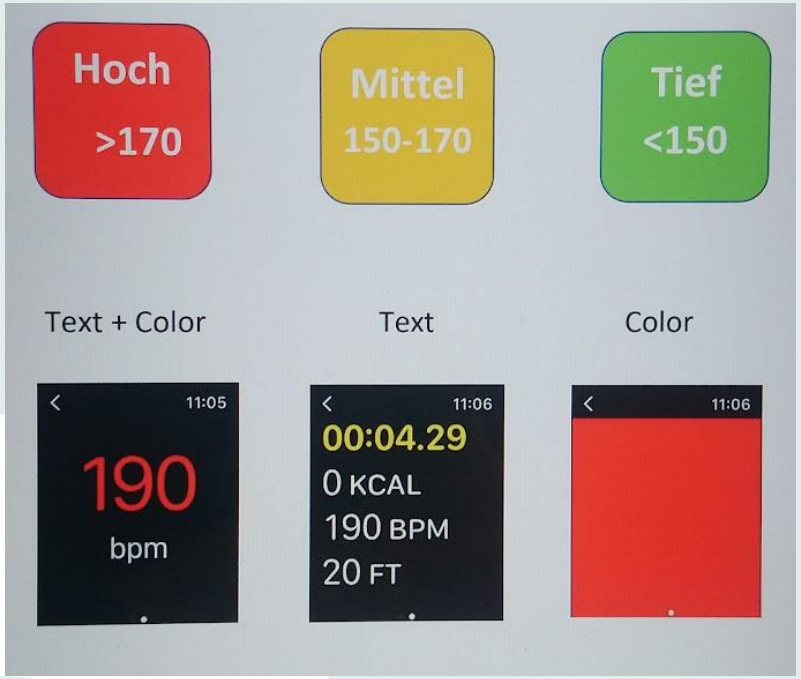
- Temperature regulated room with open windows
- Participants were asked to come in sport clothes and if they feel healthy enough to attend the study
- Treadmill to provide a safe running environment
- Ventilator that streams air to the participants whilst running on the treadmill
- Soft drinks and snacks for emergency cases like circulation problems and to provide benefits after attending the study
- Paramedic was present
- Disinfectant was used to clean



Procedure

- Greetings and Questionnaires
- Showing the participants the three distinct heart rate zones (low, moderate, high) on a tablet (see screenshot on the right)

  Not all participants were familiar on how to interpret the heart rate number - which is why a short introduction was given beforehand
- Participants set their exercise speed in order to have a “moderate” workout



Per display mode:

- Participants exercise on a treadmill
- Participants indicate heart rate zones (4x per exercise)
- Response time accuracy of the response were documented through stopwatch

3x

Procedure - Questionnaires

Beforehand:

- Privacy policy
- Demographics paper

During study (per display mode):

- Borg
- System Usability Scale

3x

6	überhaupt keine Anstrengung
7	
8	extrem locker
9	
10	sehr locker
11	locker
12	
13	ein wenig anstrengend
14	
15	anstrengend
16	
17	sehr anstrengend
18	
19	extrem anstrengend
20	maximale Anstrengung

[1] BORG- Scala, Academy of Sports, https://www.academyofsports.de/de/lexikon/borg-skala/

BORG Scale

System Usability Scale

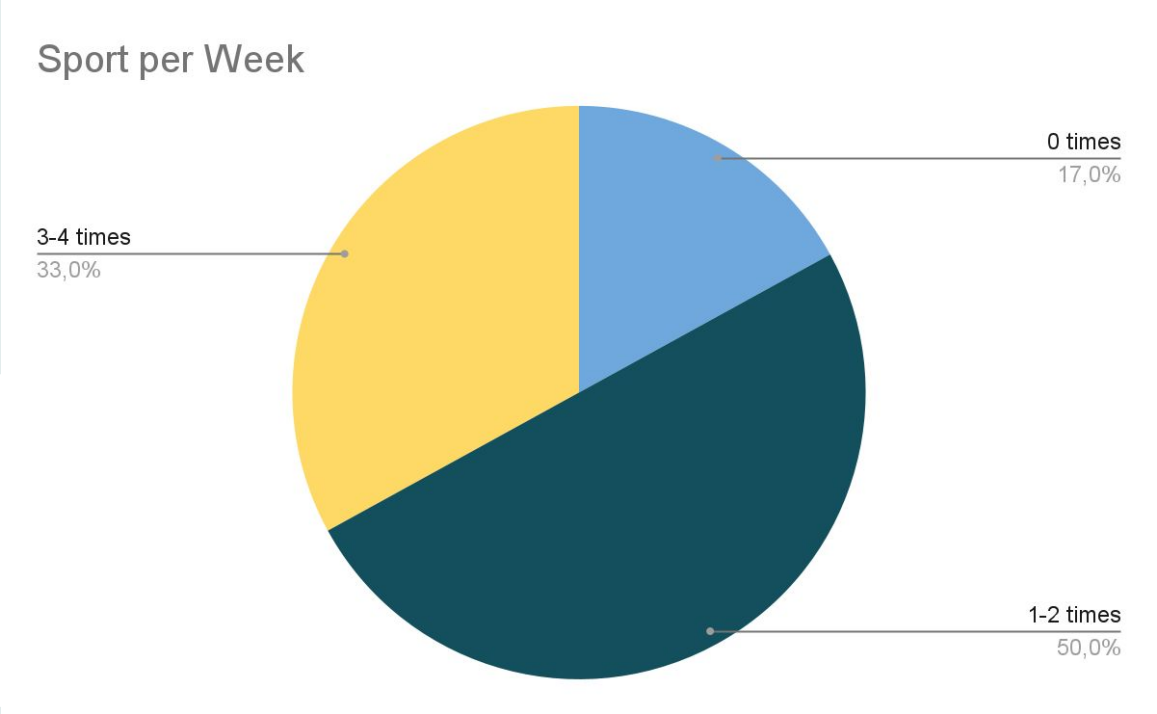
© Digital Equipment Corporation, 1986.

	Strongly disagree				Strongly agree
1. I think that I would like to use this system frequently	1	2	3	4	5
2. I found the system unnecessarily complex	1	2	3	4	5
3. I thought the system was easy to use	1	2	3	4	5
4. I think that I would need the support of a technical person to be able to use this system	1	2	3	4	5
5. I found the various functions in this system were well integrated	1	2	3	4	5
6. I thought there was too much inconsistency in this system	1	2	3	4	5
7. I would imagine that most people would learn to use this system very quickly	1	2	3	4	5
8. I found the system very cumbersome to use	1	2	3	4	5
9. I felt very confident using the system	1	2	3	4	5
10. I needed to learn a lot of things before I could get going with this system	1	2	3	4	5

System Usability Scale

Results - Summary Demographics

- Average age-range: 21-29
- 25% Male, 58% Female, 17% Diverse
- 100% Students, 91% with a Bachelor degree
- Sport per Week:
  - ➔ 0 times: 17%
  - ➔ 1 - 2 times: 50%
  - ➔ 3 - 4 times: 33%



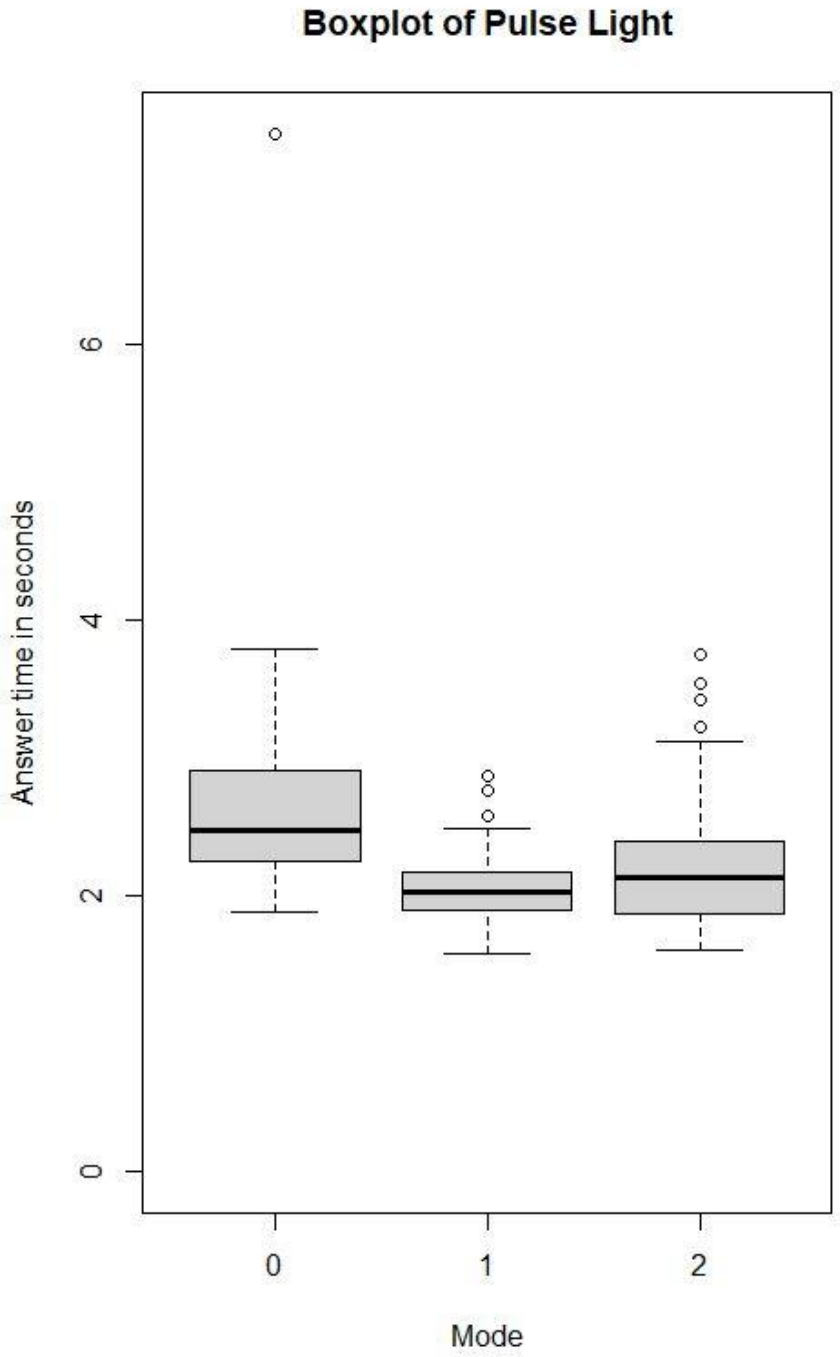
Results - Time to answer

Averages:

- Mode 0 (Text): 2.669583s
- Mode 1 (Text w. Color): 2.054792s
- Mode 2 (Color): 2.297500s

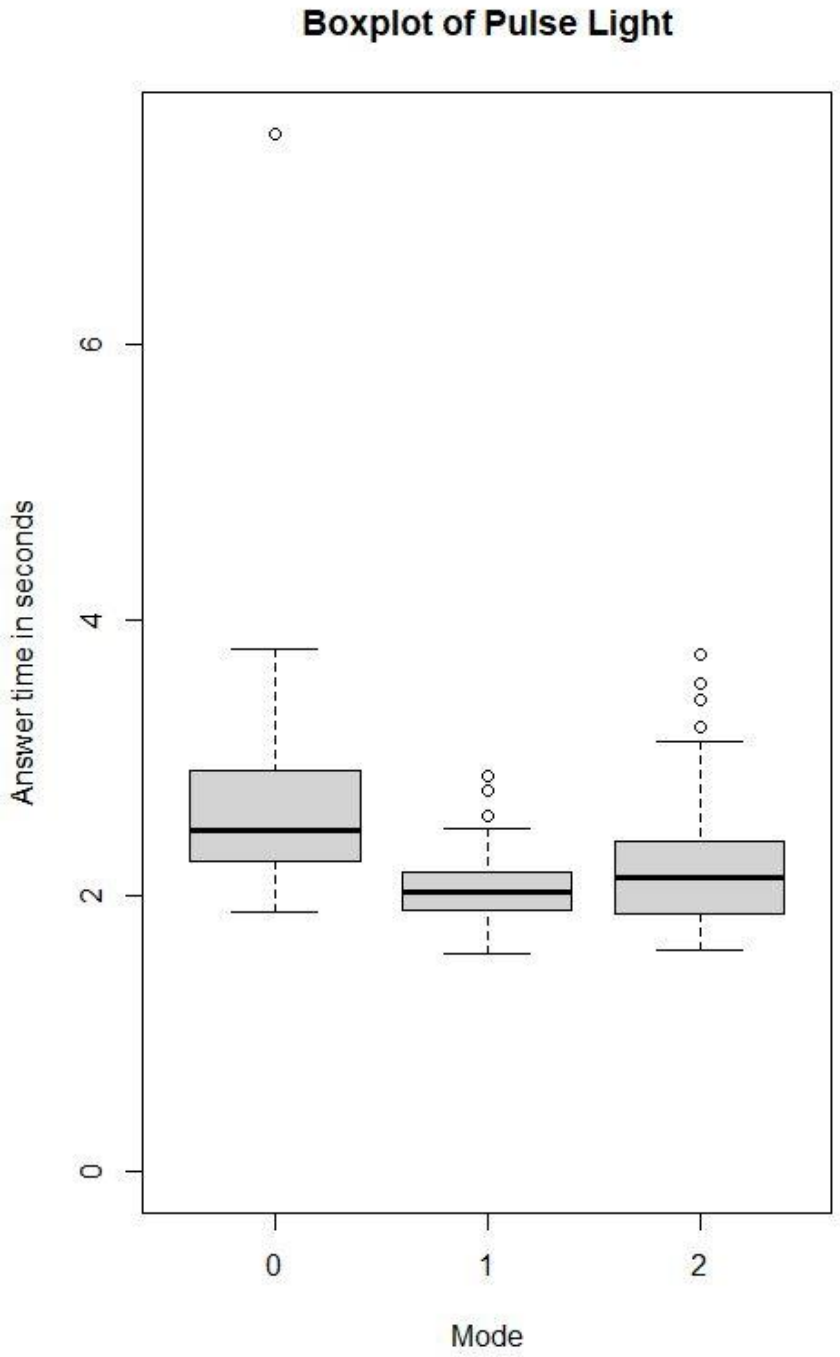
Standard Deviations:

- Mode 0 (Text): 0.8506067s
- Mode 1 (Colored Text): 0.252359s
- Mode 2 (Color): 0.5696901s



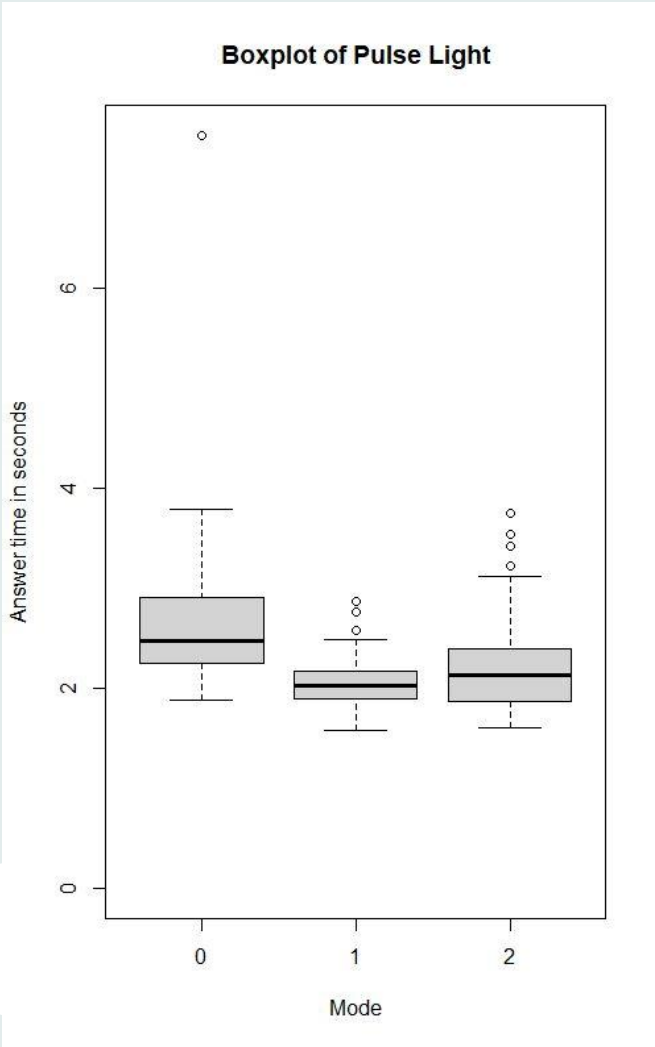
Results - Time to answer

- The data was validated with the “one-way RM ANOVA” ($p = 0.475$), showing that the difference of the three display modes is statistically significant



Results - Time to answer - Text

- “Text” (0) has not only the biggest SD (SD = 0.8506067s) but also the highest mean (M = 2.669583s) compared to the other modes
- There is one spike (7.52s) which marks the highest time-to-answer value

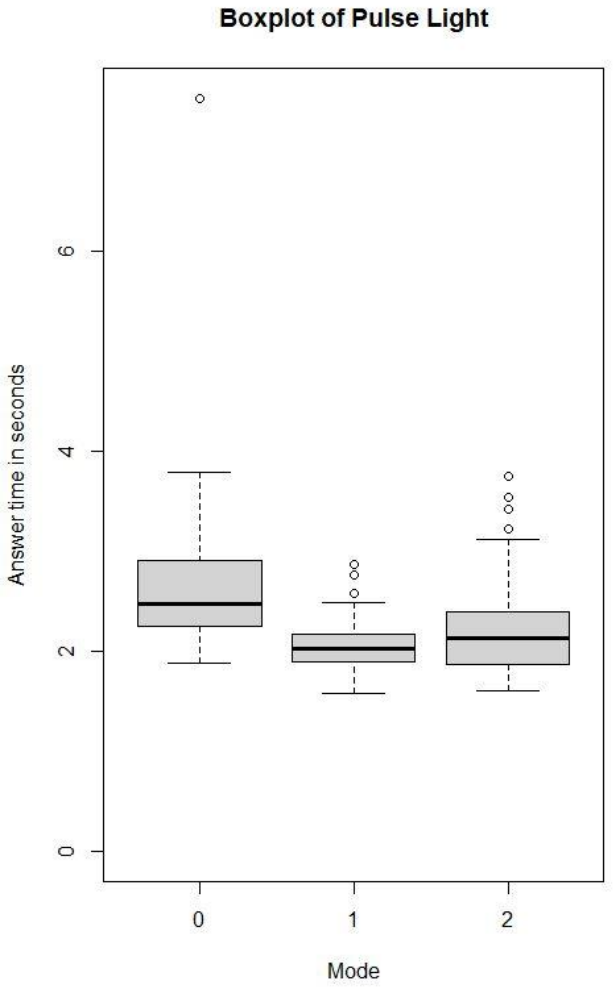


Results - Time to answer - Text w. Color

- "Text w. Color" (1) has not only the lowest SD (SD = 0.252359s) but also the lowest mean (M = 2.054792s) compared to the other modes

- With the SD of answer-value of "Text w. Color" (1) included, it is as good as the average of second placed "Color" (2) mode

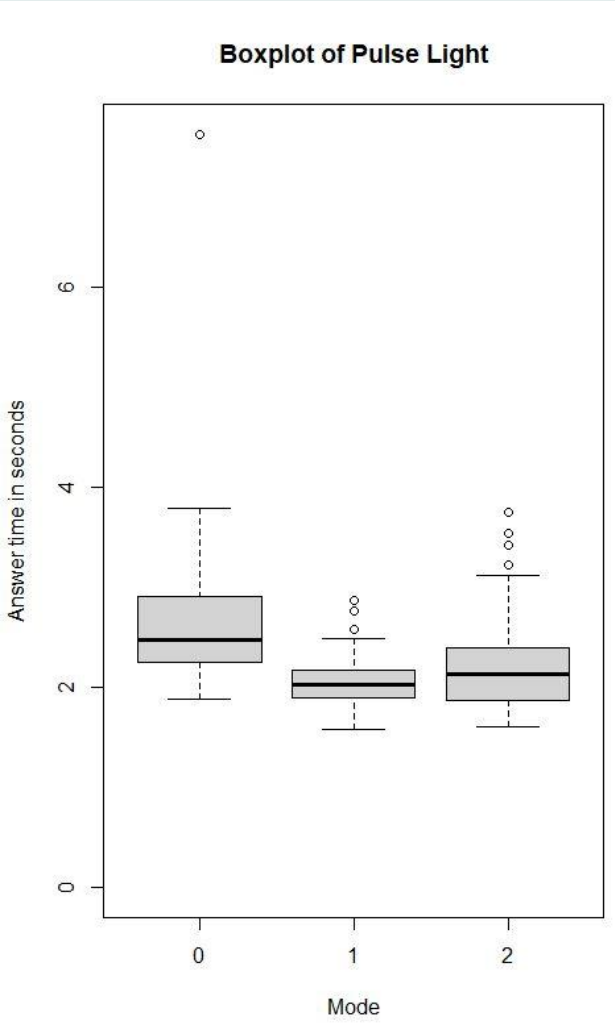
  (M_{TC}) 2.054792 + (SD_{TC}) 0.252359 = 2,307151 ~= 2.297500s (M_C)



Results - Time to answer - Color

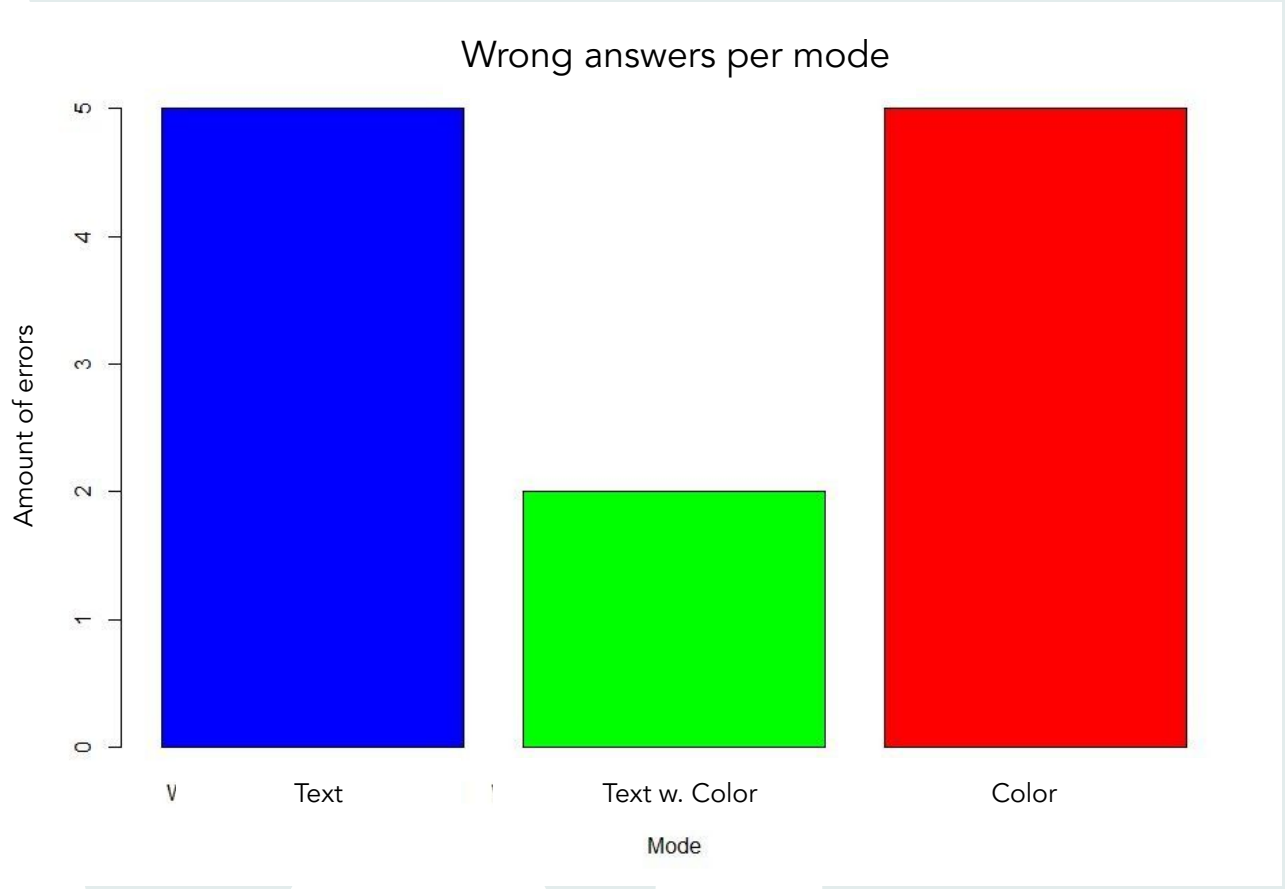
- “Text w. Color”(M= 2.054792s) and “Color” (M= 2.297500s) have similar mean values but “Color” has the greater SD value (SD = 0.5696901s) compared to “Text w. Color” (SD = 0.252359s)

- “Color” (M = 2.297500s) has a better mean value than “Text” (M = 2.669583s) with similar SD when excluding the one spike on the mode “Text”



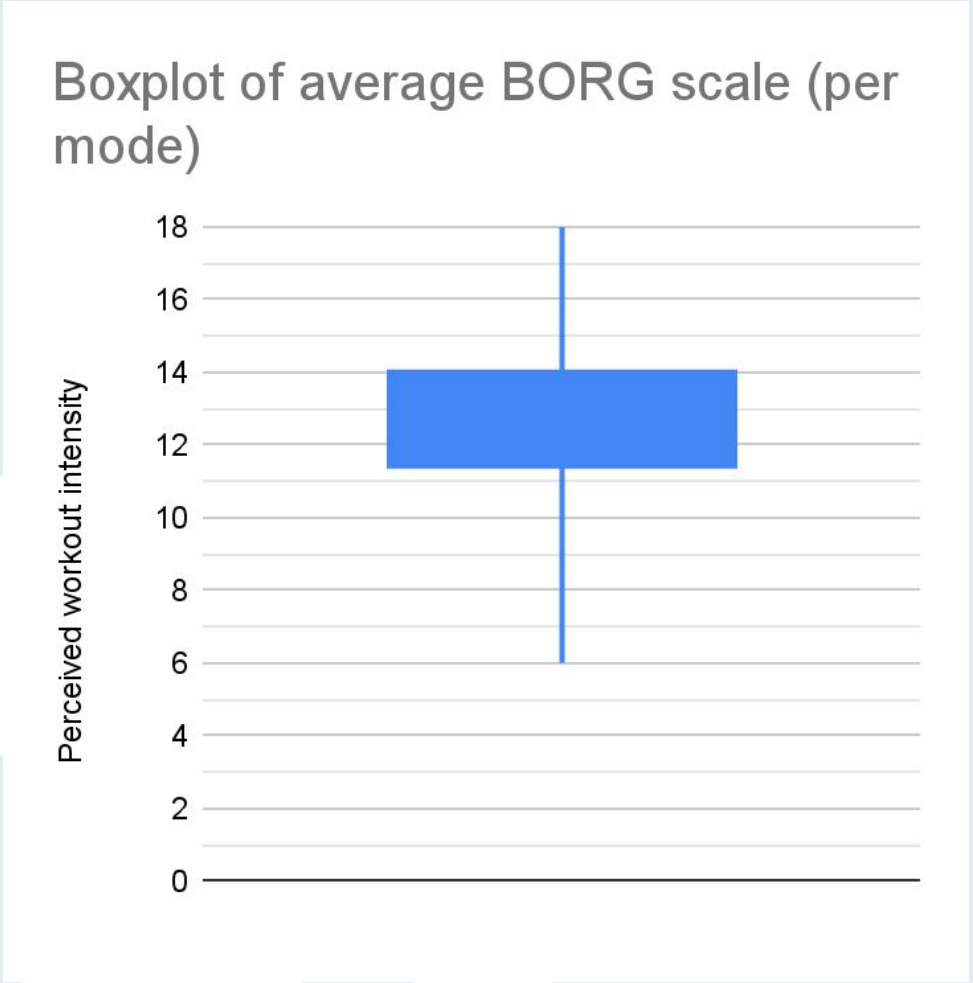
Results - Correctness of answers

- Since the correctness of the answers is either true or false (binary data), the shapiro-wilk test cannot be made
- The modes "*Text*" and "*Color*" have equal amount of errors (amount: 5)
- The mode "*Text w. Color*" has the least amount of errors (amount: 2)



Results - BORG

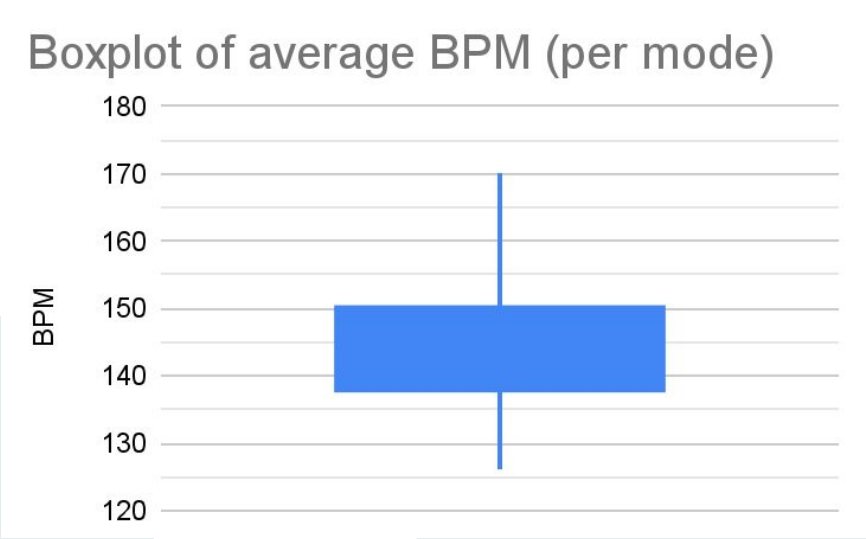
- On arithmetic average, the perceived physical intensity was 12,19
- The difference/range between highest (18) and lowest (6) is 12



Results - Correctness of BORG

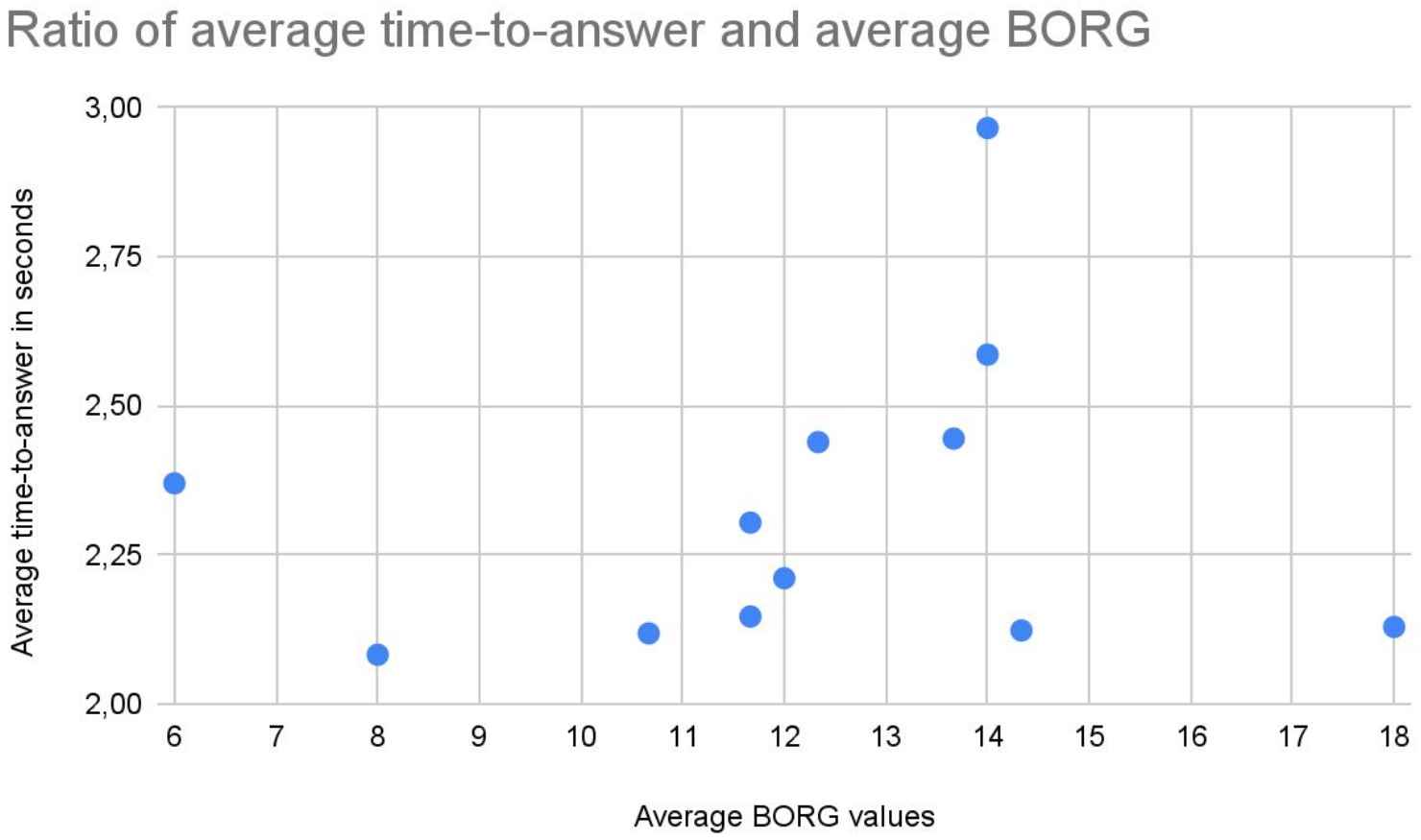
- As the BORG scale average is 12,19, through validation (and the formular BORG average = BPM average * 10), the average BPM should also be 121, yet it is 144,99 bpm
- The difference between average scale (per person and per mode) and average bpm (per person and per mode) is 23,05 beats p. minute

- In general, the actual BPM were around 2 steps higher than the perceived workout intensity
- This shows also in the difference of max (170,10) and min (126,12) = 43,977 bpm which is also roughly 20 beats p. minute



Results - Comparison BORG to Time

- As visible in the chart on the right, there seems to be no difference between the average time-to-answer values and average BORG values (both being the average of the display modes)
- If there would be a connection, there would be an increasing line with growing average time values the higher the average BORG values get



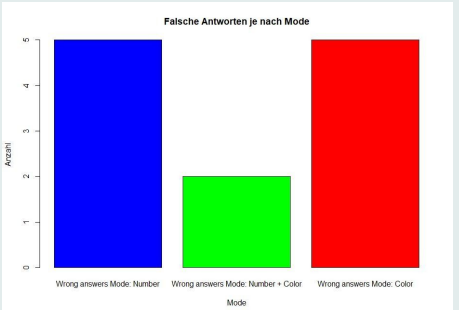
Results - SUS

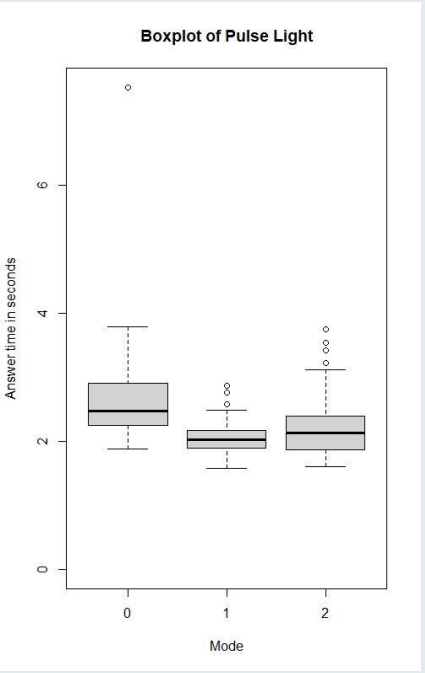
- Mode 0 (Text): 66,04166667 %
- Mode 1 (Colored Text): 71,6666666 %
- Mode 2 (Color): 68,125 %

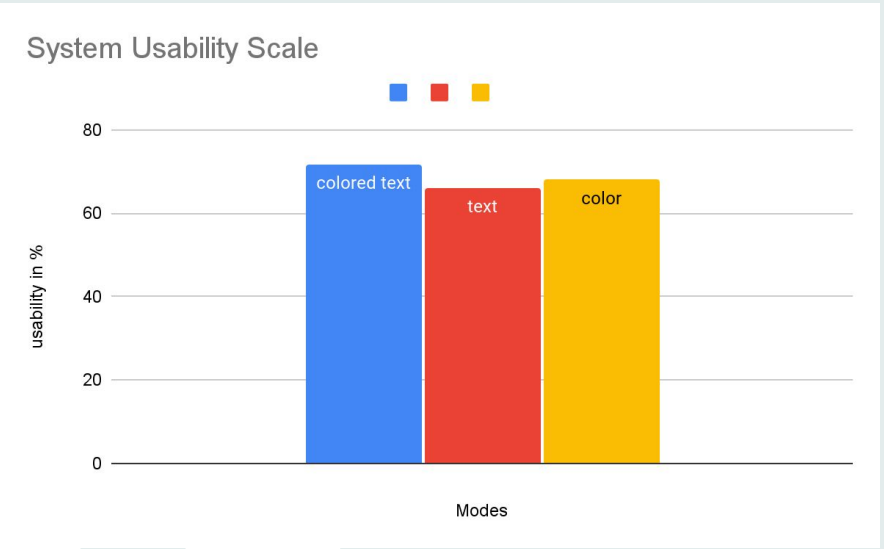
The result shows that the mode “Text w. Color” has the best usability, followed by the mode “Color” and having “Text” with the least usability. Yet the difference of all of them is quite small (~5,625 %)

Results - Mode Comparison

Text	Text with Color	Color
Lowest Sus Score 66%	Highest Sus Score 71,7%	Middle Sus Score 68,12%
Highest reaction time 2.67s	Lowest reaction time 2,05s	Middle reaction time 2,3s
Middle Borg Score 12,3	Highest Borg Score 12,5	Lowest Borg Score 11,75
Highest Error rate 5	Lowest Error rate 2	Highest Error rate 5







Implications

- Error-wise and SD-wise, the mode “Color w. Text” seems to be the most understandable
- In our study, the amount of physical activity did not have an impact on the *time to answer*
- 83% of our participants do sport regularly
- The modes “Color” and “Text” had as much errors
- “Text with Color” has a higher usability than the “Color” and “Text”

Conclusio

- The study shows that displaying one text, additionally having color, as value for BPM does have a positive impact on readability (regarding errors) and speed of comprehension

Final notes

- Participants additionally gave feedback that the Color-Only display was sometimes hard to read as some colors (green and yellow) were not greatly distinguishable when there is a fluent transition
- Additional testing after the study showed that the watch did tend to show a more green-ish color tone (e.g. when displaying yellow) when not looking perpendicular onto the display
- Participants also stated that the “Color w. Text” mode was more useful as it gave the color some black space around which gave the color some contrast and thus improve readability and distinctness
- In order to comply with health regulations, the smartwatch was disinfected after each participant

Limitations

- Due to time limitations, the study was performed with 12 participants only*
- Having used the SUS test, its results only give a fairly rough estimate whether a display mode is useable - a possible qualitative study would help to further distinguish how to improve the readability

**To further improve the study, multiple demographics groups could be relevant*

- The study excluded people with color vision deficiency