university of groningen / faculty of behavioural and social sciences / research master

Master's thesis

Reviewing current inclusive visualisation design practises within the field of psychology and investigating the efficiency of APA 7th edition compliant visualisations.

Name and initials: Daniël A. C. Vos

Student number: S2985322

E-mail address: d.a.c.vos@student.rug.nl

First assessor: Muirne C.S. Paap & Casper J. Albers

Second assessor: Christian E.G. Steglich

Programme: Research Master Behavioural and Social Sciences

Theme: Understanding Societal Change

ECs: 30

Date: 27-3-2023

Word count: 6700

Are there deviations of the Master's thesis from the proposed plan?
☒No
☐Yes, please explain below the deviations

1.1 Abstract

The recent release of the APA 7th edition includes a section about colour-use in scientific publications to make data visualisations inclusive for readers with a visual deficiency. The purpose of this study is twofold: (1) assess the state of the implementation of the new colour-use guidelines in a small sample of APA journals in 2020 – 2021, and (2) determine how the new colour-use guidelines affect non-visual deficient readers. The assessment of the data visualisations in the APA journals follows a systematic review approach, with a grading instrument based on the APA colour-use guidelines and their considered best-practice. The effectiveness of the inclusive data visualisation for non-visually deficient readers follows a survey-based experimental design in which participants received either an APA compliant or non-compliant data visualisation and are scored on their answer accuracy and self-perceived mental load. The results from the systematic review indicate that the implementation can be improved in both colour-use and application of redundant coding practices, and the results from the survey-based experimental design show no difference in efficiency between an APA compliant and non-compliant data visualisation. Which indicates that current data visualisation practices can be improved and there is no barrier for implementing more inclusive data visualisation practices.

Keywords: Data visualisation, Colour vision deficiency, APA, Colour-use, Colour contrast, WCAG guidelines, Redundant coding, Psychology, Social Science, Statistics.

1.2 Reviewing current inclusive visualisation design practices within the field of psychology and investigating the efficiency of APA 7th edition compliant visualisations.

When an author decides to place a data visualisation within their research paper, authors face several design decisions, which may impact the interpretability of that data visualisation. One of the most impactful design decisions is the implementation of colour. This design decision is why the APA included a separate section in the 7th edition on colour use in data visualisations, which expands the previous data visualisation guidelines in other editions. This addition to the existing guidelines aligns with the current colour usage trend in scientific publications due to the transition from print to digital publishing (Cole, 2004; Davidson, 2005; Iliinsky & Steele, 2010). The main benefit of implementing colour within a data visualisation is that it enables the designer to add a layer of information over the data visualisation without hindering the interpretation process of the reader due to the colour being a pre-attentive process (Iliinsky & Steele, 2010). This process means that colours require no effort to differentiate from each other due to the evolution of our visual system. Even though colours require little effort to read for people with normal colour vision, the colour used in data visualisation can impact the readability and interpretation of people who experience some form of visual deficiency such as colour vision deficiency (CVD) (Crameri, Shephard, & Heron, 2020; Stone, Laskowski, & Lowry, 2008). CVD is one of the most common conditions in the human visual system, impacting 1 out of every 12 men (8%) and 1 out of every 200 women (0.5%) (Simunovic, 2010). Strikingly, little is known about the current colour-use practices within psychology when designing data visualisations and their effects on readers with CVD.

CVD is the inability to distinguish certain shades of colour (Simunovic, 2010). Distinguishing colours is made possible by the communication between the brain and the photoreceptor cones in the retina. The human visual system is trichromatic, meaning that each type of photoreceptor cone is active when translating light to perceivable colours. The three types can be summarised as red, green, and blue photoreceptor cones, with each type possessing its distinctive wavelength of light can perceive (figure 1, appendix A). However, when the light processing of one or more of the three cones becomes aberrant or deficient, the trichromatic visual system becomes colour vision deficient (Simunovic, 2010).
CVD comes in several levels of severity, with the most extreme variation being the rarest. The least severe form of colour vision deficiency is *anomalous trichromacy*, which implies that all three cones are active; however, how light is processed is aberrant. Therefore, the perceivable colour can differ from the object's actual colour (Simunovic, 2010). A more severe form of colour vision deficiency is *dichromacy*, meaning one out of three cones is inactive. In such cases, the wavelengths to be processed by the inactive photoreceptor cone remain unprocessed, resulting in an inability to perceive that particular colour and its variants. Dichromatic vision is most prevalent for the red and green cones (Simunovic, 2010). The most severe and rarest form of colour vision deficiency is *monochromacy*, which has only one active cone resulting in a colour view composed of shades of grey or blue. Importantly, even though colour and the saturation of those colours are impossible for readers with CVD to differentiate, the luminance level is still perceivable even for people with *monochromacy* (Stone & Laskowski, 2008).

Luminance is one out of three inherent properties of a colour. The first property is hue, which is the attribute of the colour that makes it red, green, blue, yellow, or any intermediate between a contiguous pair of these colours (Edwards, 2004). Saturation is the second attribute, which can be described as the purest hue of colour without dilution with white (Edwards, 2004). However, depending on the severity of the CVD, the visibility of the hue and saturation property is limited. The property that CVD does not limit is luminance, or in other words, the perceived lightness of a colour. Whereas saturation is the purest form of a colour, luminance is that pure colour diluted by white, and this dilution is perceivable by almost every reader with CVD (Edwards, 2004; Stone et al., 2008).

The difference between different luminance levels is known as colour contrast, which can be expressed in a ratio. This ratio is calculated by dividing the relative luminance of the lighter and darker colours, resulting in a ratio from 1:1 to 21:1 (WCAG 2.0, 2016). The APA 7th ed. colour use guidelines state that the WCAG 2.0 AA standard for non-textual objects is the best practice. The AA compliance level from the WCAG 2.0 states that each coloured data element (lines, dots, bars, and areas) requires a colour contrast ratio of at least 3:1 with its background (see fig 2 for an example of a line, appendix A). As seen in figure 2, the visualisation with the 3:1 contrast ratio distinguishes itself better from the background than the 1.7:1 colour contrast visualisation. When more than one coloured data element is present within the visualisation, the colours should both contrast with their background and each other (WCAG 2.0, 2016)

The figure shown by the APA to demonstrate their guidelines compliant with the WCAG 2.0 success criterion for non-textual objects did surprisingly not pass the minimally sufficient colour contrast standard. Additionally, the colours within the visualisation are derived from a non-scientific colour pallet, which are pallets that are known to distort the data, unintuitive to interpret, exclude and discriminate against people with any form of CVD, and not readable when printed to black and white (Crameri et al., 2020). Whilst scientific pallets are perceptually uniform, perceptually ordered, CVD friendly, translatable to black and white, and are peer-reviewed (Crameri, 2018; Crameri et al., 2020). The example given by the APA shows how difficult it is to apply the WCAG 2.0 in practice, which can result in an erroneous adaptation by authors and publishers using the APA as their paper formatting guidelines. Therefore, a review of the current colour use practises and their level of compliance within APA journals should result in an accurate overview of the data visualisation design areas in which authors and journal editors can improve.

The only attempt at studying inclusive data visualisation design within psychology publications has been published by Frane (2015) who studied colour used in the scientific journals *Psychological Science, Current Directions in Psychological Science,* and *Clinical Psychological Science* between September 2013 and August 2014. From all the studied data visualisations, 8.1% were considered confusing or unreadable by a multidisciplinary panel consisting of both colour vision deficient and normal colour vision academics from various fields. However, the wide range of possible severity levels of colour vision deficiency and including only readers with dichromacy could have resulted in underestimating the number of problematic visualisations. I argue that a more objective metric can yield better results.

With the release of the APA 7th edition colour use guidelines, a testable objective metric for problematic colour use now exists the 3:1 colour contrast ratio. This ratio has two benefits over subjective judging of graphs: (1) the metric is not tailored for a specific form of colour vision deficiency, and (2) it can be calculated by software made available by the WCAG, which is considered best practice by the APA. In addition to the colour contrast ratio, the APA colour use guidelines state that including redundant coding can remedy possible problematic colour use within publications. Redundant coding adds a layer to the existing data visualisation to enhance the readability and separability of coloured data elements (Wilke, 2019). The most commonly used redundant coding techniques are glyphic symbols, different textures, and using labels instead of legends (see figure 3 for an example, appendix A). However, redundant coding is a point of contention within data visualisation design. Some proponents actively advocate using redundant coding since it results in more functional and effective means to display complex relationships within a data visualisation (Bateman et al., 2010; Hullman, Adar, & Shah, 2011). On the opposing end are the minimalists, who actively try to minimise the clutter within a data visualisation, which they argue benefits interpretability since everything on display has statistical meaning (Blasio & Bisantz, 2002; Gillan & Richman, 1994; Tufte, 2006). According to Vanderplas & Hofmann (2020), the evidence suggests that redundant coding enhances data visualisations and represents a departure from Tufte's "above all else, show the data" principles. However, more studies are required to fully understand the effect of redundant coding. Therefore, this study will test whether the implementation of minimally sufficient contrast and redundant coding impact the efficiency of normal-visioned readers since there is little research on how normal-visioned readers experience these changes in data visualisation designs.

The study draws on the cognitive load theoretical framework to measure the effectiveness of the APA 7th edition compliant data visualisations. This theoretical framework is commonly used to understand the cognitive demands of a specific design. Cognitive load theory assumes that learners have two forms of memory: long-term and working memory. Long-term memory is where information is stored for later use and is drawn upon when necessary (Sweller, Ayres, & Kalyuga, 2011). In the long-term memory resides our data visualisation schema, which allows us to recognise the data visualisation format and assists the interpretation and comprehension by using experiences from the past (Shah & Hoeffner, 2002). The working memory is utilised when we perform a task, such as using the data visualisation schema to answer questions about the data visualisation. However, unlike long-term memory, working memory is finite and can store a maximum of four pieces of information simultaneously (Cowan, 2001). When the maximum amount of working memory is exceeded, the learner experiences a cognitive overload that impairs the ability to comprehend the information displayed (Sweller et al., 2011).

The cognitive load theory states that cognitive load can take two shapes: intrinsic and extraneous (Paas & Van Merrienboer, 1993; Sweller et al., 2011). Intrinsic load is user-centric, meaning that the experienced load is regulated by the user's knowledge about the subject, which varies per user and is out of the control of the data visualisation designer (Sweller et al., 2011). Extraneous load is designer-centric since this form of load encompasses how the information is displayed, and a visualisation designer should try to minimise this particular load (Sweller et al., 2011). This begs the question of whether the inclusion of redundant coding and high colour contrast, as described in the APA 7th edition colour-use guidelines, will minimise or increase the extraneous cognitive load.

Paas and Merrienboer (1993)developed a metric to measure extraneous load by estimating the relative educational efficiency. It comprises two distinct features: accuracy (thus, how well the user answered the problem) and self-reported mental effort (how much mental effort it took to complete the task). For instance, high accuracy and a low mental effort signify that the extraneous load is limited. In contrast, high accuracy with a high mental load means that the data visualisations have a substantial extraneous load. The relative efficiency metric function summarises both the accuracy and mental load, with an efficiency score. A score of 0 means that the visualisation is balanced in both accuracy and mental load, and any score above 0 implies an imbalance in either accuracy or self-perceived mental load. The relative educational efficiency allows us to compare two data visualisations with either an APA colour-use guideline-compliant format versus a format that uses a minimalistic design without redundant coding.

Not every aesthetic is weighted the same when it comes to cognitive load due to the existence of pre-attentive processes. Colour is one of those pre-attentive processes, meaning that recognising colours acts parallel to the working memory and occupies minimal space in the working memory (Iliinsky & Steele, 2010). Nevertheless, research has shown that colour can affect the interpretability of data visualisations since colour enables the user to separate groups and other objects quickly. Therefore, the same educational efficiency metric is calculated for data visualisations with greyscale colour coding to discover whether the addition of colour coding in data visualisation yields better efficiency scores or that the need for colour increases when the complexity of the data visualisation increases.

In summary, with the release of the APA 7th edition colour use guidelines, a step has been made towards offering more accessible data visualisations for readers with a colour vision deficiency. However, it remains unknown how authors publishing in APA journals adopt these new design guidelines. In this study, we will address the adaptation of the APA 7th colour-use guidelines by answering the following research question:

RQ1: To what degree do articles published after the release of the APA 7th edition complies with the new colour use guidelines used within psychological research?

Furthermore, little research exists on how normal colour vision readers will experience these new design guidelines in practice when encountering them in APA journals, which leads to the final two research questions and hypotheses:

RQ2: To what extent do WCAG 2.0 compliant figures affect the cognitive load experienced by readers without colour vision deficiency?
H1: Participants who received the APA 7th compliant data visualisations will score higher on accuracy than those who received the non-compliant alternative.
H2: The APA 7th compliant data visualisations will score lower on self-perceived mental load when compared to the non-compliant variant.

As a final research question, readers' preference for either coloured or greyscale visualisations is inquired. Since the APA still actively encourages greyscale data visualisations and colours should only increase the interpretability when the data visualisation increases in complexity (APA style, 2022; Iliinsky & Steele, 2010; Wilke, 2019). Which results in the following final research question and hypothesis:

RQ3: To what extent does figure complexity influence the reader's preference for colour over grayscale?
H2: The participants will prefer the greyscale data visualisations for the simple variant and the coloured variant when the data visualisation increases in complexity.

2 Methods

To address all research questions, the paper is divided into two studies that will be addressed separately. RQ1 will be addressed by the first study using a systematic review approach. The other study following a experimental design approach addresses RQ2 & RQ3.

2.1 Study 1: Systematic review of published data visualisations within APA journals

2.1.1 Eligibility criteria for inclusion in the systematic review

For the systematic review, the following journals are selected Psychological Science, Clinical Psychological Science, and Current Directions in Psychological Science between January 2020 and December 2021. These journals were chosen due to (1) offering a wide range of different subjects within the field of psychology, and (2) all journals are APA curated. The sampling period is between January 2020 and December 2021 since the APA 7th edition was released in October 2019.

From the selected journals, every published paper that included a coloured data visualisation was eligible to be included. From the selected papers, the first coloured figure was extracted as long as the visualisation displayed some form of statistical data, such as (line graphs, bar plots, boxplots, and maps). Therefore schematic representations of a model (such as path diagrams); EEG scans, or pictures are not

2.1.2 Evaluation Procedure

Each data visualisation included in the current study was labelled as either compliant, semi-compliant, or non-compliant. The compliant and non-compliant classifications were based on the WCAG 2.0 standards of non-textual objects in online environments in conjunction with the guidelines published by the APA. We chose to add the semi-compliant category since the guidelines are stringent, and we wanted to separate visualisations that used minimally sufficient colour contrast from visualisations that did not. Every data visualisation was evaluated using the colour contrast analyser (CCA) developed per the WCAG 2.0 standards. This tool enables us to translate colours to hexagonal values and calculate the minimally sufficient contrast. To come to a decision, two evaluations took place with every data visualisation, with the first being the evaluation of the colour contrast and the other being the use of contrasting aesthetics. A singular reviewer does both evaluations.

2.1.3 Determining the level of compliance

To be labelled compliant with the APA colour-use guidelines, the data visualisations should pass the 3:1 colour contrast threshold for every included colour with the background. If there is an overlap between coloured data elements, the colours should either have a 3:1 colour ratio or include contrasting aesthetics, enabling the reader to separate both data elements.

Semi-compliant data visualisation has colours that have a 3:1 colour contrast threshold with the background. However, it cannot be considered fully compliant due to colour contrast issues between the overlapping and bordering data elements.

Lastly, non-compliant data visualisation possesses at least one colour which does not pass the 3:1 colour threshold with its background. Therefore, it can be considered illegible for individuals with a visual impairment.

2.2 Study 2: Assessing the efficiency differences between APA-compliant and non-compliant visualisations.

2.2.1 Participants

For the current study, a convenience sample of 52 participants was recruited using posters, flyers, and face-to-face recruitment. The recruitment process took place at the Faculty of Behavioural and Social Sciences of the University of Groningen, PPO (postdoctorale psychologie opleidingen), utilising contacts from my professional network. The participant had to meet the following inclusion criteria: (1) be an undergraduate student or a graduate student (2) have at least passed one university-level statistics course.

The Ethics Committee Psychology of the University of Groningen has approved this study (PED-2122-S-0077)

2.2.2 Procedure

The study used Qualtrics for a survey-based experiment where participants could fill in on any device with an internet browser (Qualtrics, 2023). During the survey, the participants could skip questions or return to previous questions, and there was no time limit. The median completion time of the survey was 9.45 minutes.

Besides questions about the participant's colour vision and whether or not they passed a university-level statistics course, no other background variables were collected. The reasoning behind not collecting any additional background variables was that it is difficult to generalise results to a population using a convenience sample.

Each participant was randomly allocated to one of the two conditions comprising four data visualisations, each following an identical structure. The first two data visualisations had three questions, each featuring line graphs in colour and greyscale, known for their easy readability and interpretation. The final two data visualisations were scatterplots presented in colour and greyscale. The scatterplot is considered more complex due to the number of elements (increased difficulty in legibility due to various colours and shapes) and the type of questions asked (requires statistical knowledge). The conditions were either getting data visualisations compliant with the APA 7th edition colour use guidelines, whilst the others were not. The data visualisation compliant with the APA 7th edition colour-use guidelines used colours with at least a 3:1 colour contrast ratio with their respective background and redundant coding to emphasise differences between the data elements. The non-compliant counterpart used the default colour palette of ggplot2 (v3.4.1; (Wickham, 2016), which is known not to pass the 3:1 colour contrast ratio and is widely used as a default colour palette (Long & Teetor, 2019). Furthermore, the non-compliant visualisation did not use redundant coding to emphasise differences. The entire survey can be found in Appendix C

The structure for both conditions was identical, starting with a brief introduction about the soon-to-be-presented data visualisation. After the introduction, the data visualisation was displayed, and three task-related questions were displayed on the same page. All the tasks were designed such that it was necessary to differentiate the data elements from each other. The tasks which the participants were asked to complete were the following:

Q1. Report the average value of a particular group at a specific point in time.

Q2. Report a specific group's lower- and upper-bound ranges across all time points.

Q3. Report the perceived trend of a group based on given directions of a trend.

After being presented with the three task-related questions, the participant was asked to rate the mental effort spent on the three tasks using the 10-point mental effort scale developed by Paas (1992). After answering the questions for both data visualisation's colour variations, the participant was asked to indicate their preference for either the greyscale or the coloured variant. Below this question was an open text field without a word limit, which the participant could use to elaborate on their reasoning for this preference. This process was repeated for the greyscale alternative, which was identical for both groups.

For the second round of data visualisations, a similar structure was used. It started with a short introduction, followed by three task-related questions, measuring their self-reported mental effort spent to complete the three tasks and closing the section with the preference question. Following the preference question, the participants received a three-variable scatterplot, which followed a similar structure to the line graph. The tasks that were displayed below the scatterplot were the following:
Q1. Identify a trend for a specific group.
Q2. Estimate the correlation between the dependent variables for a specific group.
Q3. Identify the type of relationship a specific group has with the dependent variables.

2.2.3 Measures

2.2.3.1 Task accuracy

Task accuracy was calculated by summing the number of times a participant answered a question correctly, followed by a division by three (see formula below).

$$(1)\ Task\ accuracy = \frac{(Q_1 + Q_2 + Q_3)}{3}$$

Table 1 (appendix B) shows the answers scored as correct for the tasks within the survey. For questions where the point floats above a round integer, a range of correct was accepted since it would be too strict about taking only one exact correct answer. Answers ranging from 5.2 were still scored with a 0.5 instead of 1 if they were within the accepted range of answers.

2.2.3.2 Self-reported mental effort

The current study uses the mental effort scale developed by Paas, van Merrienboer, & Adam (1994). The instrument comprises of one question "*How much mental effort did it cost to solve the three questions?*" for which the participant is presented a 9 point Likert scale with answers ranging from 1 (very, very low mental effort) to 9 (very, very high mental effort).

Instructional efficiency is calculated by combining task accuracy and self-perceived mental effort (Paas & Van Merrienboer, 1993). First, by transforming both metrics to z-scores and then taking their absolute sum, followed by dividing the combined score by the square root of the two (see the formula below). You divide by the square root of two to reduce the variance back to one.

$$(2)\ Instructional\ Efficiency = \frac{\left|z_{Task\ accuracy} - z_{Mental\ effort}\right|}{\sqrt{2}}$$

2.2.4 Analysis plan

The purpose of this study is to conduct a two-way repeated measures ANOVA to examine the effects of APA compliance (between factors) on the instructional efficiency (dependent variable) on varying levels of data visualisation task complexity (within factor). Before conducting the analysis, the data were checked for missing values and possible outliers. The model assumptions were checked to ensure the results were valid and reliable. The data should be approximately normally distributed using a Shapiro-Wilk test. For the homogeneity

of variances, Levene's tests is used. For the final assumption of sphericity, Mauchly's Test of Sphericity is used; in case the sphericity assumption is violated, a Greenhouse-Geisser correction would be applied to adjust the degrees of freedom for the repeated measures ANOVA.
The software package R (R Core Team, 2022) was used for two-way repeated measures ANOVA.

2.2.5 Power and sample size

To determine the sample size an a priori power analysis was run using GPower (v3.1; (Erdfelder, Faul, & Buchner, 1996) for a two-way RM-ANOVA with the following specifications: medium effect size (f = .25), with an error probability of .05 and a power of .95. Resulting in a minimal sample size of 54 participants. Unfortunately, the study did not meet the required power level by 2 respondents. However, this should not impact the credibility of the findings.

2.3 Results

2.4 Study 1: Results from the systematic review of published data visualisations within APA journals.

In 2020 and 2021, a total of 575 articles were published in Psychological Science (275), Clinical Psychological Science (145), and Current Directions in Psychological Science (155). Of these articles, 273 contained at least one coloured data element (line, bar, boxplot, dot), which comprises 47.4% of all articles. From these coloured data visualisations, 57.9% used the rainbow/jet colour palette. From all the collected data visualisations, 22.3% used only one colour, which can be considered as a purely aesthetic feature instead of a functional one.

Table 2 (appendix B) displays data visualisation types studied across all three journals from 2020 to 2021. The publications in 2020 show that the most popular data visualisation format is either a line plot or a bar plot. A similar trend can be identified in 2021, with line plots and bar plots used most frequently. The other category is the third most used classification in 2020 and 2021, which is understandable due to it being a category which collects all visualisations that did not fit in the other classifications (such as heatmaps, alluvial plots, and 3D plots).

Comparing the compliance levels of the visualisations used in the studied publications across the three journals (figure 4 (appendix A), table 3 (appendix B)), however, psychological science stands out with its proportion of non-compliant data visualisations within their publications over 2020 and 2021. The proportion of non-compliant data visualisations is notably lower in Clinical Psychological Science and Current Directions Psychological Science. However, the association between the degree of non-compliance and the journal is not statistically significant ($X^2 = 7.19$, df = 4, $p = 0.13$). Furthermore, one could expect non-compliance to decrease over the years by getting accustomed to the APA 7th edition guidelines since every journal sees a decrease in non-compliant visualisations. However, there was no statistically significant association between the year of publication and the level of compliance ($X^2 = 0.42$, df = 2, $p = 0.81$).

Contrasting aesthetics are rarely used across the journals (Figure 5, appendix A). The majority of published data visualisations used no contrasting aesthetics. For the remainder of the visualisations that did use contrasting aesthetics, different textures and glyphic symbols were the most commonly used. In-graph labelling or a combination of contrasting aesthetics were rare occurrences in the studied publications.

2.5 Study 2: Results from assessing differences in interpretation efficiency for APA compliant and non-compliant data visualisations.

2.5.1 Descriptive statistics

The current study contained 73 participants from mixed nationalities who are either studying psychology or are psychology graduates. 22 participants were excluded from the analysis due to either being colour-vision deficient, not agreeing with using their data, or not having passed a university-taught statistics course. After inspection of the given answers, no missing values were identified. Each participant that met the inclusion criteria was randomly allocated to either the compliant (*n* = 30) or non-compliant (*n* = 22) visualisation group. The exclusion of the participants who did not meet the inclusion criteria resulted in unequal condition groups due to how Qualtrics allocates the participants. The way Qualtrics allocates the respondents were identified after the data collection was concluded; therefore, it could not be remedied. Tables 4 and 5 (appendix B) show the summary statistics for participants in both conditions on each level of complexity for both colour variants. When comparing the participants' accuracy on the simple data visualisation task, the participants presented with the non-compliant version performed better than the compliant version ($t(49.99) = -2.36,\ adj.p = 0.022$). However, this difference does not extend to the complex data visualisation task. The only notable difference in accuracy is that for the non-compliant group, no participant got more than one correct answer; still, this did not result in a substantial mean difference between the compliant and non-compliant visualisation groups.

Regarding mental load, there were minimal differences between the compliant and non-compliant data visualisation groups for both the simple and complex variants. These minimal differences also explain why the compliant and non-compliant data visualisation are close to each other regarding efficiency (see Figures 6 & 7, appendix A). There are a couple of outliers for the complex compliant data visualisations; however, the results are nearly identical for most other answers. A similar plot is made for the greyscale alternative (figure 8), showing similar patterns and no increased efficiency in simple or complex data visualisation.

When comparing the coloured data visualisations with the greyscale alternatives, there were some notable differences. For the simple data visualisation, the overall accuracy of the greyscale alternative is significantly lower than its coloured counterpart ($t(97.27) = -6.04$, $p < .001$), with a significantly higher required mental load ($t(101.98) = 3.25,\ p = .003$). Surprisingly, this result is the opposite for the complex alternative, which has a significantly higher average accuracy than its coloured alternative ($t(88.30) = 12.55,\ p < .001$). Nevertheless, the mental load still is significantly higher for the greyscale alternative when compared to its coloured counterpart ($t(101.92) = 2.72,\ p = .016$).

2.5.2 Qualitative results

Even though there are no differences in efficiency, there are substantial differences in user preference when asked to choose between coloured or greyscale. The overwhelming

“Colours make the differentiation between groups way easier as does the label of the group directly besides the lines. The grayscale variant is harder to read, specifically between the standing and treadmill desk.” [14]
“It’s easier to distinguish the groups and I had to look at the legend fewer times” [8]
“Easier to distinguish between groups with the colours and the dots help” [25]

majority for the simple data visualisation indicated preferring the coloured (80%) compared to 13% for the greyscale variant. The difference is even more striking for the complex variant, for which every participant preferred the coloured variant (100%) over the greyscale. For both visualisations, the primary motivation for preferencing the coloured over the greyscale alternative was that, according to the participants, the lines and dots in coloured variants were more straightforward to distinguish than in the greyscale alternative. Below are some quotes for participants that received the compliant data visualisation.
A similar pattern was found when looking at the preference for non-compliant data visualisation, as seen below.

Similar motivations were given for the complex data visualisation task, such as easier to distinguish, and the greyscale with just the glyphic symbols took more effort to interpret. Thus, even though the participants perceived the coloured visualisation as the better option, the greyscale variant resulted in overall higher accuracy. This contradicts the results in tables 2 and 3 (appendix B), where the mental load is slightly higher than their coloured counterparts; however, the overall accuracy is also higher.

2.6 Model results

Table 6 (appendix B) shows the results from the multiple RM-ANOVA models. One of the most

> "*I can distinguish the three different colours rather easily, but distinguishing the two "dotted lines" is slightly more difficult*." [17]
> "*Difference between groups is more clear. In the grayscale, standing desk and treadmill look too similar*." [20]
> "*Colour difference is much more obvious to distinguish than grayscale variant*" [54]

notable results is that the level of compliance did not result in significant mean differences in accuracy, mental effort, and efficiency. Complexity, however, significantly affected the participant's overall accuracy when completing the tasks. Furthermore, the interaction between compliance and complexity is only significant for the accuracy and not the self-perceived mental effort. However, this interaction between compliance and complexity is insignificant for the overall data visualisation efficiency.

2.6.1 Model assumptions

The two-way repeated measures ANOVA has multiple assumptions that must be met. The first is that each model's dependent variables should be measured on either continuous level, which is the case for every estimated model. Furthermore is, the within-subjects factor measured with two categorical levels. There are no significant outliers in any combination of the groups. The normality assumption, however, is problematic since Shapiro tests yield violations for this assumption for every model. Fortunately, the two-way repeated measures ANOVA is robust against violations of the assumptions. Finally, every model passes Mauchly's test of sphericity.

2.7 Discussion

The key findings from both the systematic review of data visualisation practices within three APA-curated journals and the experimental design that studied the effectiveness of APA-compliant data visualisations for readers without CVD are the following: (1) the majority of published data visualisations do not comply with the APA 7th edition best practice; (2)

redundant coding is rarely used in any of the published data visualisations; (3) using redundant coding can reduce the interpretation accuracy for readers without CVD; (4) there exists no difference in efficiency between APA compliant and non-compliant data visualisations.

Within psychology, there is little research to compare the results from the systematic review with, apart from the previously discussed Frane (2015) study. Nevertheless, a similar statistic is available from a source outside of the field of psychology. For instance, Chartability (2022a), a workbook that aids data visualiser in creating inclusive data visualisations, published the proportion of audits that failed due to lack of colour contrast (88%). This is similar to the results from the systematic review when taking both the non-compliant and semi-compliant visualisations together. A possible explanation for this similarity in results is that the steps required to pass the colour contrast criterion are relatively easy to check. However, it requires knowledge of what to compare (Chartability, 2022b). This leads, to the first recommendation for the APA colour use guidelines, since the guidelines are unclear on how to approach WCAG 2.0 compliant colour contrast. The APA only mentions that a figure should have 'plenty' of colour contrast and that the colour contrast analyser can assist with calculating the contrast. However, the APA fails to mention between which elements the colour contrast is required. Therefore, the APA should clarify how to approach appropriate colour contrast and offer a step-by-step implementation guide according to the WCAG 2.0 AA success criterion for non-textual objects.

Another strategy to comply with the APA colour-use guidelines is the implementation of redundant coding. However, across all three journals, only a few publications used some form of redundant coding in conjunction with colour. Even though including redundant coding is stated by the APA to enhance the data visualisation, the experimental results yield contradictory results as displayed in the results from the experimental design. The accuracy for readers that received data visualisations with redundant coding performed less on the accuracy metric than those that received data visualisations without redundant coding. A possible explanation for the lack of accuracy is that the different textures used within the visualisation are similar in that they have different combinations of dashes and dots, which some participants labelled as confusing. Therefore, the APA should refrain from stating that redundant coding is a solution for non-compliant colour contrast but should offer some additional guidelines on how and when to apply redundant coding more in-depth.

Taking all things together, there is a solid foundation within the APA guidelines to build upon. Currently, the guidelines do not impede readers' efficiency in interpreting data visualisations; therefore, journals shouldn't enforce them more rigorously to create a scientific landscape that is inclusive for readers with CVD.

2.7.1 Limitations and future research

Due to the current study being a master's thesis, it did not get any additional funding, which resulted in only one reviewer doing the entire systematic review. As a result, some threats to the systematic reviews' conclusion validity got introduced—possible subjective coding bias when judging the published visualisations. However, the constructs chosen for the review are nearly free of bias, since they use a subjective metric, the colour contrast ratio, and the other components, such as the inclusion of redundant coding, leave little room to add subjectivity within the coding process. Even though the effect is minimal, a future study should use a second reviewer to code the same visualisation to circumvent this issue simultaneously.

Another limitation within the systematic review concerns the construct validity of the colour contrast calculations. To elaborate, what is meant by this is that due to technical limitations, no source code or hex codes were available to the reviewer to extract the 'true' colours within the data visualisation. Therefore, the reviewer used the colour droplet in the

CCA tool, which enables the reviewer to pick a pixel and extract the hex code from that particular pixel. However, not every pixel has the same colour, which could introduce variation in the CCA outputs hex code. Even though this variation can present different results when picking different pixels, the reviewer always picked the centre or middle pixel of a coloured object to prevent or at least contain the variation in hex codes received. For future research, it would prove beneficial if the reviewer got the source code of the data visualisation, which enables the reviewer to extract the true colour, which is not impacted by variations introduced by the droplet feature of the CCA tool.

The key limitation within the experimental design concern the construct validity of the used self-perceived mental effort scale. The way this scale is implemented in the current experimental design only measures the mental effort using one question, which can impact the validity of the metric. Even though Paas et al. (1994) claim that a singular question yields similar results to a metric comprised of multiple questions. As an alternative to adding additional questions, adding time elements could improve the validity of this particular metric, done in a similar study by Huang, Eades, & Hong (2009), who studied an extraneous load of different social network graphs. Within this study, the authors combined the self-perceived mental effort scale and the time element in the efficiency calculation. Therefore, additional measurements for the mental effort scale are beneficial for future studies and should be pursued if the design allows it.

Another limitation of the experimental design concerns the internal validity of the design; the results attained within the experiment cannot and should not be translated to a broader population since it is attained using a convenience sampling method. In the initial proposal for the research, the target population was students within the field of psychology; however, due to difficulties in attaining a sample of sufficient size, the inclusion criteria got broadened. Participants were no longer required to be a student and could, for instance, already have a position within the field (e.g. practitioner or lecturer). As a result, making claims about the population of psychology students became impossible. Therefore the results should indicate no difference in efficiency between compliant and non-compliant data visualisations but cannot be translated to a particular population within the field.

The final limitation of the experimental design concerns the conclusion validity due to heavily skewed groups within the design. Within the analysis, the compliant group got substantially more participants than the non-compliant group, which is attributed to how Qualtrics allocates participants. To elaborate, every time a participant accepted the informed consent form Qualtrics would allocate the participant to either the compliant or the non-compliant group while trying to achieve a fifty-fifty split. However, since not everyone completed the entire survey (which resulted in the deletion of the response), Qualtrics still perceived the incomplete response a complete response, which is the cause of the skewness. Unfortunately, there is no easy-to-implement solution within Qualtrics; therefore, perhaps using different software that allows more flexibility would benefit future research.

References

APA style. (2022). Accessible use of color in figures. Retrieved from https://apastyle.apa.org/style-grammar-guidelines/tables-figures/colors#:~:text=Authors%20seeking%20publication%20should%20avoid,color%20reproduction%20for%20printed%20materials.

Bateman, S., Mandryk, R., Gutwin, C., Genest, A., McDine, D., & Brooks, C. (Apr 10, 2010). Useful junk? Paper presented at the 2573-2582. doi:10.1145/1753326.1753716 Retrieved from http://dl.acm.org/citation.cfm?id=1753716

Blasio, A. J., & Bisantz, A. M. (2002). A comparison of the effects of data–ink ratio on performance with dynamic displays in a monitoring task. *International Journal of Industrial Ergonomics, 30*(2), 89-101. doi:10.1016/S0169-8141(02)00074-4

Chartability. (2022a). Chartability. Retrieved from https://chartability.fizz.studio/

Chartability. (2022b). The chartability workbook. Retrieved from https://chartability.github.io/POUR-CAF/

Cole, B. L. (2004). The handicap of abnormal colour vision. *Clinical and Experimental Optometry, 87*(4-5), 258-275. doi:10.1111/j.1444-0938.2004.tb05056.x

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *The Behavioral and Brain Sciences, 24*(1), 87-114. doi:10.1017/S0140525X01003922

Crameri, F. (2018). Geodynamic diagnostics, scientific visualisation and StagLab 3.0. *Geoscientific Model Development, 11*(6), 2541-2562. doi:10.5194/gmd-11-2541-2018

Crameri, F., Shephard, G. E., & Heron, P. J. (2020). The misuse of colour in science communication. *Nature Communications, 11*(1), 5444. doi:10.1038/s41467-020-19160-7

Davidson, L. A. (2005). The end of print: Digitization and its Consequence—Revolutionary changes in scholarly and social communication and in scientific research. *International Journal of Toxicology, 24*(1), 25-34. doi:10.1080/10915810590921351

Edwards, B. (2004). *Color*. East Rutherford: Penguin Publishing Group. Retrieved from https://ebookcentral.proquest.com/lib/[SITE_ID]/detail.action?docID=6060279

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers, 28*(1), 1-11. doi:10.3758/BF03203630

Frane, A. (2015). A call for considering color vision deficiency when creating graphics for psychology reports. *The Journal of General Psychology, 142*(3), 194-211. doi:10.1080/00221309.2015.1063475

Gillan, D. J., & Richman, E. H. (1994). Minimalism and the syntax of graphs. *Human Factors, 36*(4), 619-644. doi:10.1177/001872089403600405

Huang, W., Eades, P., & Hong, S. (2009). Measuring effectiveness of graph visualizations: A cognitive load perspective. *Information Visualization, 8*(3), 139-152. doi:10.1057/ivs.2009.10

Hullman, J., Adar, E., & Shah, P. (2011). Benefitting InfoVis with visual difficulties. *IEEE Transactions on Visualization and Computer Graphics, 17*(12), 2213-2222. doi:10.1109/TVCG.2011.175

Iliinsky, N., & Steele, J. (2010). *Beautiful visualization: Looking at data through the eyes of experts* O'Reilly. Retrieved from http://cds.cern.ch/record/1385925

Long, J. D., & Teetor, P. (2019). *R cookbook: Proven recipes for data analysis, statistics, and graphics* O'Reilly Media. Retrieved from http://cds.cern.ch/record/2682819

Paas, F. G. W. C., & Van Merrienboer, J. J. G. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors, 35*(4), 737-743. doi:10.1177/001872089303500412

Paas, F. G. W. C., van Merrienboer, J. J. G., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills, 79*(1), 419-430. doi:10.2466/pms.1994.79.1.419

Qualtrics. (2023). Qualtrics. Retrieved from https://www.qualtrics.com

R Core Team. (2022). R: A language and environment for statistical computing [computer software]. Vienna, Austria:

Shah, P., & Hoeffner, J. (2002). Review of graph comprehension research: Implications for instruction. *Educational Psychology Review, 14*(1), 47-69. doi:10.1023/A:1013180410169

Simunovic, M. P. (2010). Colour vision deficiency. *Eye (London, England), 24*(5), 747–755. doi:https://doi.org/10.1038/eye.2009.251

Stone, M., Laskowski, S. J., & Lowry, S. Z. (2008). *Guidelines for using color in voting systems*. Gaithersburg, MD: U.S. Dept. of Commerce, National Institute of Standards and Technology.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory* (1st ed.). New York [u.a.]: Springer. doi:10.1007/978-1-4419-8126-4 Retrieved from http://bvbr.bib-bvb.de:8991/F?func=service&doc_library=BVB01&local_base=BVB01&doc_number=021195810&sequence=000004&line_number=0002&func_code=DB_RECORDS&service_type=MEDIA

Tufte, E. R. (2006). *The visual display of quantitative information* (2. ed., [Nachdr.] ed.). Cheshire, Conn: Graphics Press. Retrieved from http://www.gbv.de/dms/ilmenau/toc/516508245.PDF

Vanderplas, S., Cook, D., & Hofmann, H. (2020). Testing statistical charts: What makes a good graph? *Annual Review of Statistics and its Application, 7*(1), 61-88. doi:10.1146/annurev-statistics-031219-041252

WCAG 2.0. (2016). Understanding SC 1.4.3
. Retrieved from https://www.w3.org/TR/UNDERSTANDING-WCAG20/visual-audio-contrast-contrast.html

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis* (2nd ed. 2016 ed.). Cham: Springer International Publishing. doi:10.1007/978-3-319-24277-4 Retrieved from https://library.biblioboard.com/viewer/32b393a0-c3d0-11ea-a9fb-0ae0aa0d175d

Wilke, C. O. (2019). *Fundamentals of data visualization*. Sebastopol: O'Reilly Media, Incorporated. Retrieved from https://ebookcentral.proquest.com/lib/[SITE_ID]/detail.action?docID=5734202

3 Appendix A – Figures

Figure 1

depicts the light sensitivity of every cone type in the human trichromatic visual system. with blue having the lowest sensitivity, Green, and red overlapping. Data acquired from Stockman & Sharpe (2000).

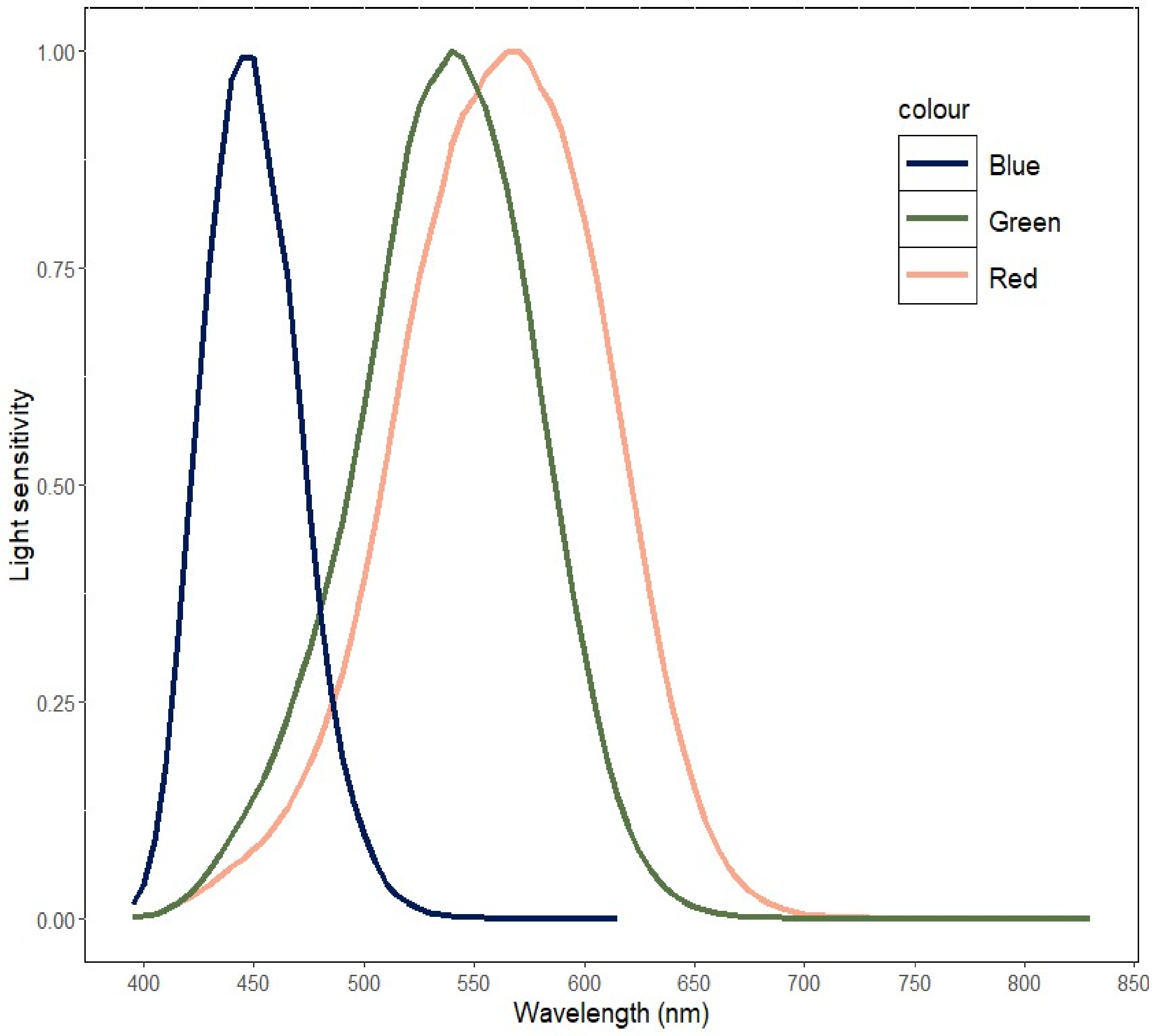


Figure 2

An example of both a non-compliant colour contrast ratio (left) and a compliant contrast ratio (right). Both examples have identical hue's and saturation levels, though vary on their level of luminance, with left high luminance and right low luminance.

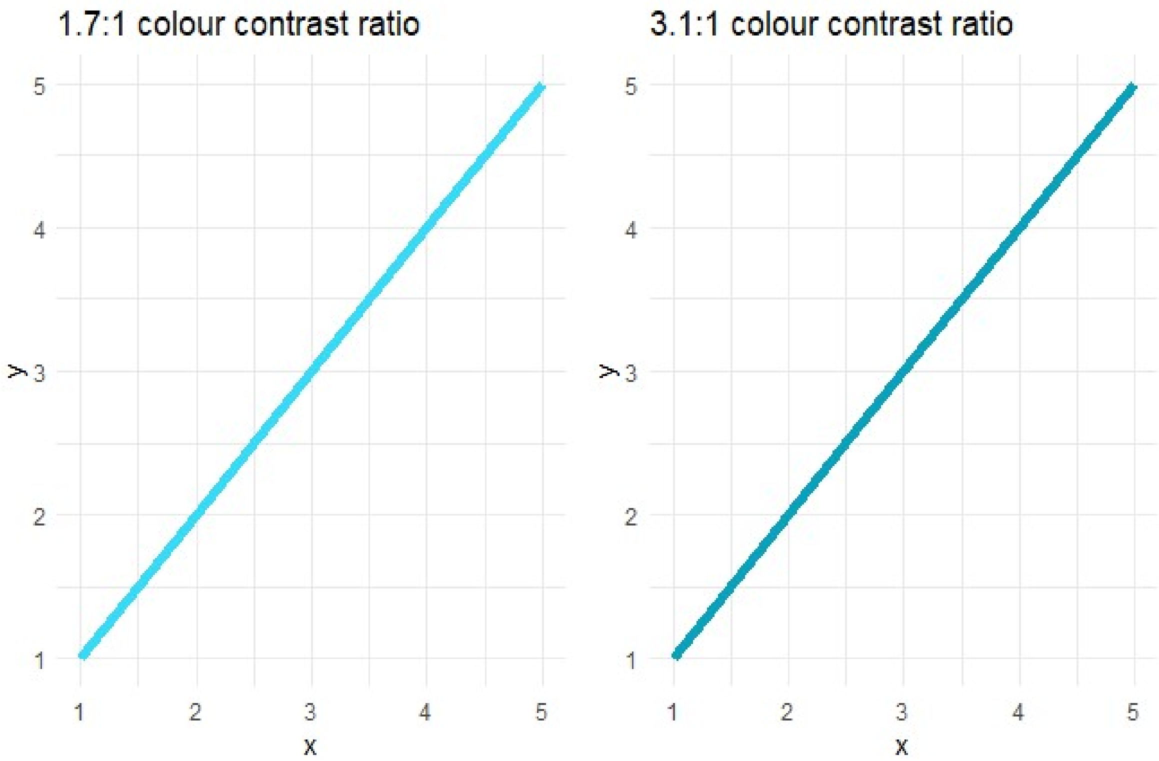


Figure 3

Examples of redundant coding with on the left a coloured data visualisation without redundant coding (left) and the redundant coded alternative (right).

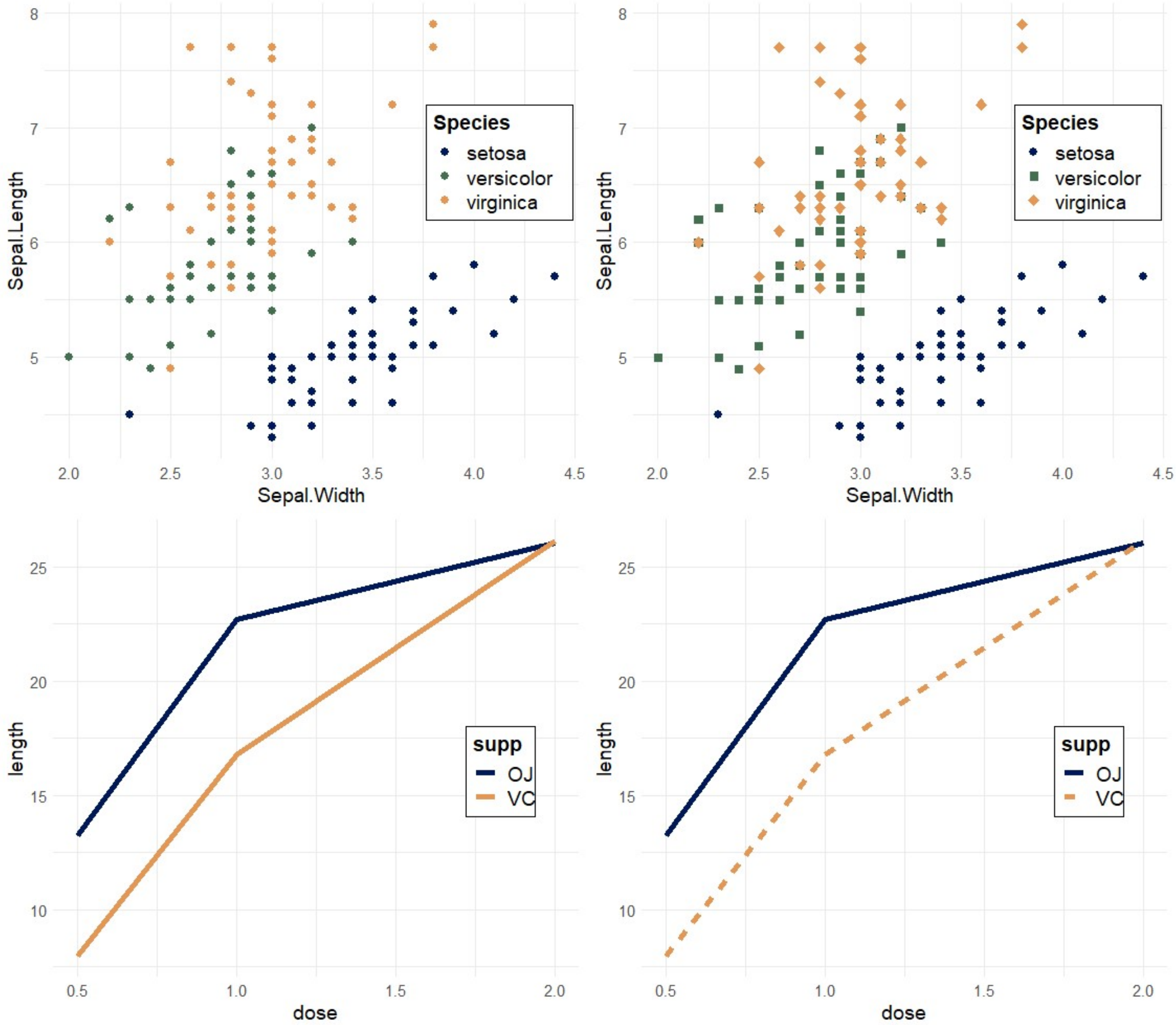


Figure 4

Proportions of the level of compliance across Psychological Science, Clinical Psychological, and Current Directions in Psychological Science.

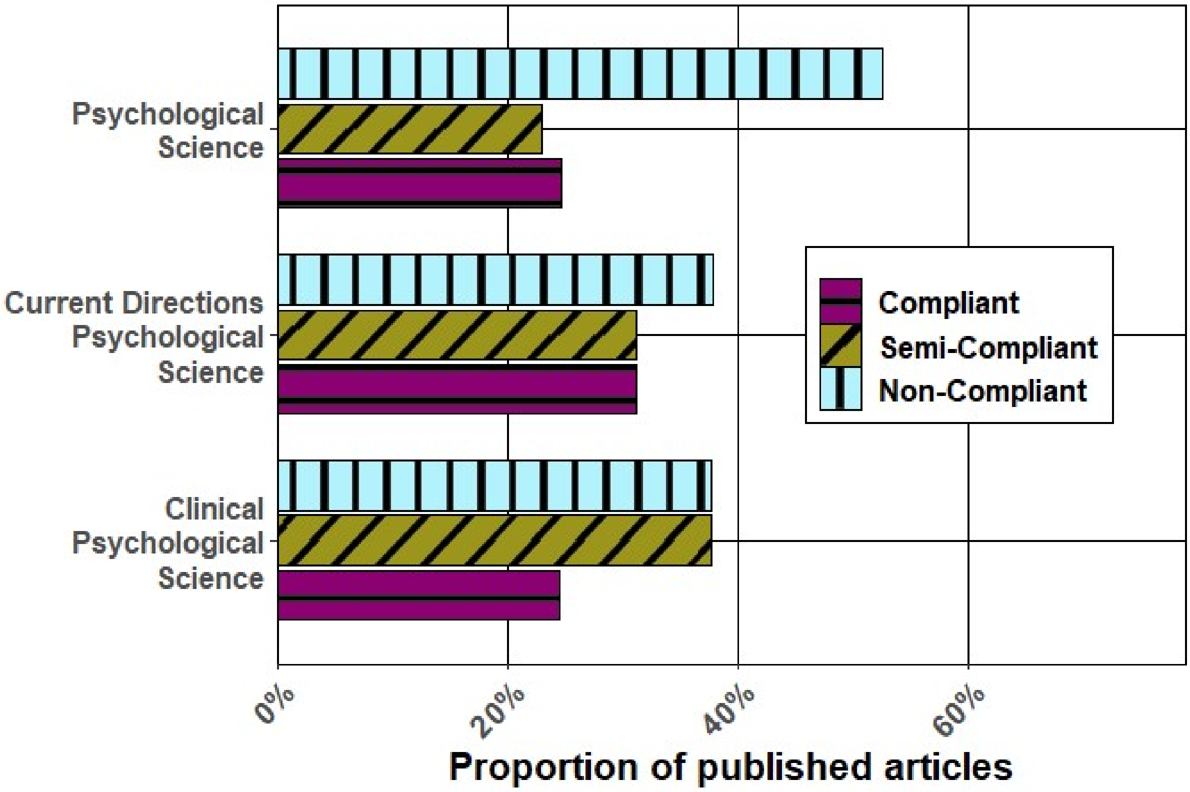


Figure 5

The proportion of data visualisations that use a particular form of redundant coding or a combination of multiple methods.

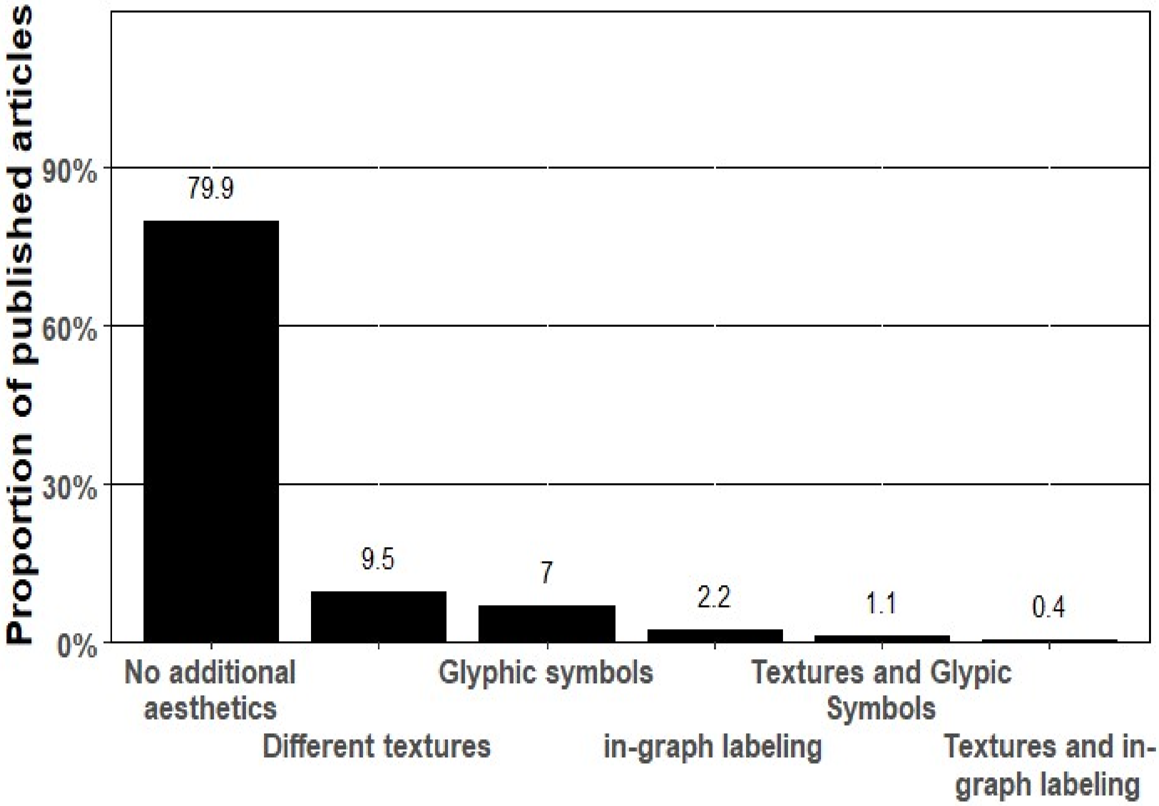


Figure 6

Efficiency scores of the participants for both the compliant and non-compliant data visualisations for each complexity level

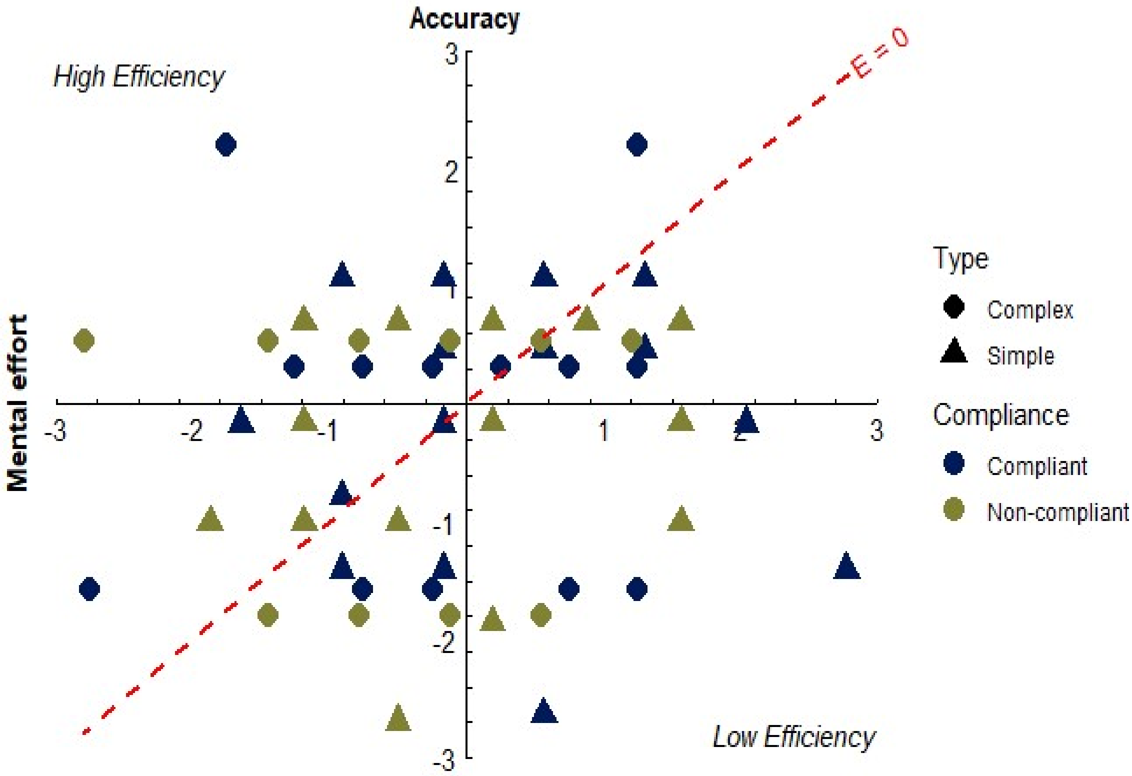
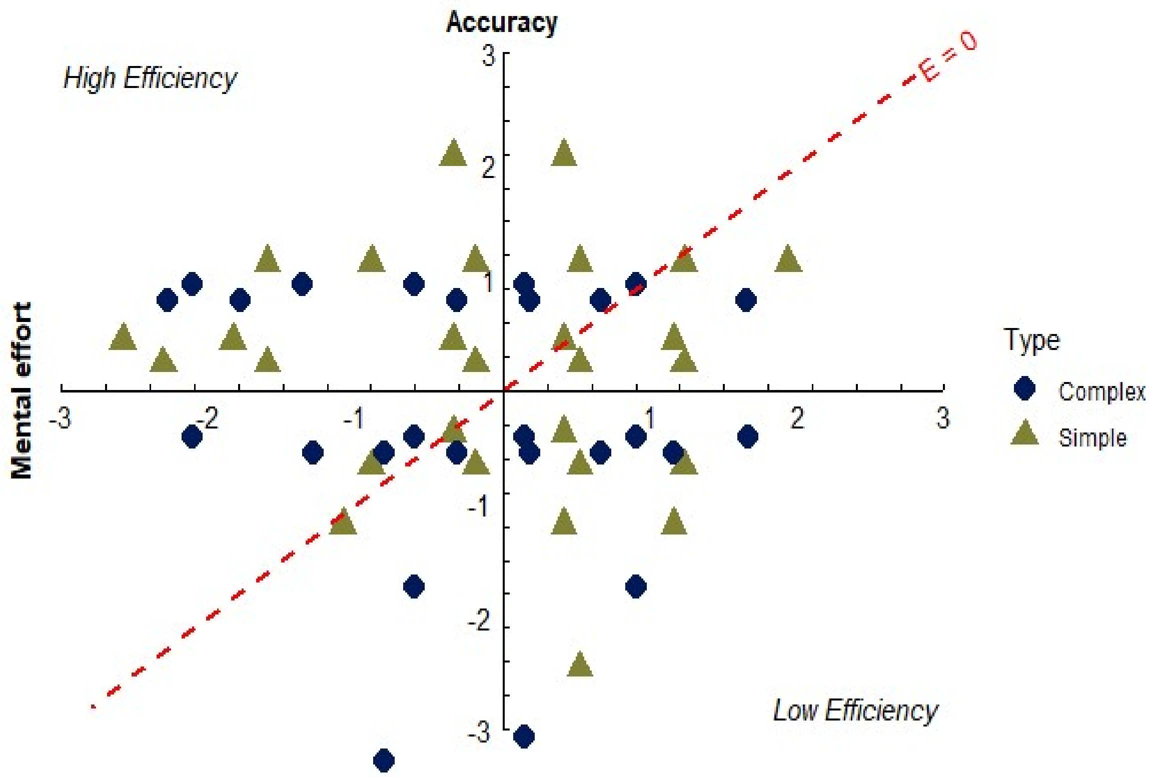


Figure 7

Efficiency scores of the participants for greyscale data visualisations for each complexity level



3.1 Appendix B – Tables

Table 1

A tabular summary of the correct answers within the survey

	Coloured visualisation		Greyscale visualisation	
	Correct answer	Accepted range	Correct answer	Accepted Range
		Line graph*		
Q1	5.20	[5.10:5.30]	5.80	[5.70;5.90]
Q2	[3.00;7.00]		[4.00;6.00]	
Q3	Standing Desk		Normal Desk	
		Scatter plot*		
Q1	Negative		Negative	
Q2	Between .75 and 1.00		Between .00 and .25	
Q3	Non-linear		Negative-linear	

* The actual phrasing of the questions can be found in Appendix A

Table 2

Proportions for the types of data visualisations judged in Psychological Science, Clinical Psychological Science, and Current Directions in Psychological Science, based on the first coloured data visualisation from each studied paper.

	2020		2021	
	N	%	N	%
Line plot	51	41.4%	52	34.6%
Box plot	4	3.2%	13	8.6%
Scatter plot	10	8.1%	19	12.6%
Bar plot	35	28.4%	44	29.3%
Other	23	18.6%	22	14.6%
Total	123	100%	150	100%

Note. This table portrays the distribution of the most commonly used graphical formats in the studied publications. Most notably, the line plot and bar plot are popular within the published articles.

Table 3

Level of compliance per journal separated by year of publication.

Level of compliance	2020 N	2020 %	2021 N	2021 %
	Clinical Psychological Science			
Compliant	6	22.2%	7	26.9%
Semi-compliant	11	40.7%	9	34.6%
Non-compliant	10	37.1%	10	26.9%
	Current Directions Psychological Science			
Compliant	7	30.4%	7	31.8%
Semi-compliant	8	34.8%	6	40.9%
Non-compliant	8	34.8%	9	31.8%
	Psychological Science			
Compliant	19	26%	24	31.8%
Semi-compliant	12	16.4%	28	27.3%
Non-compliant	42	57.5%	50	40.9%

Note. This table displays both the proportional distribution of the different compliance levels for each studied journal. The number of visualisations in Psychological science is substantially higher, since it had double the issues of the other journals.

Table 4

Summary statistics for the compliant and non-compliant coloured visualisations

	Compliant (n = 30)				Non-compliant (n = 22)			
	Mean (SD)	Median	Min	Max	Mean (SD)	Median	Min	Max
	Simple interpretation tasks							
Accuracy	0.71 (0.27)	0.83	.00	1.00	.86 (.20)	1.00	.33	1.00
Mental Load	4.23 (1.36)	4.00	2.00	8.00	3.73 (1.45)	3.50	1.00	6.00
Efficiency*	.72 (.68)	.52	.01	2.95	.74 (.48)	.68	.13	1.81
	Complex interpretation tasks							
Accuracy	.28 (.18)	.33	.00	.67	.26 (.14)	.33	.00	.33
Mental Load	5.50 (2.00)	6.00	.00	8.00	5.18 (1.50)	6.00	1.00	7.00
Efficiency*	.74 (.72)	.62	.04	2.79	.64 (.66)	.47	.01	2.35

* Based on the Paas and Meriënboer (1992) equation.

Table 5

Summary statistics for the greyscale variant.

	Simple interpretation tasks (n = 52)			
	Mean (SD)	Median	Min	Max
Accuracy	.51 .20	.50	.00	1.00
Mental Load	4.92 (1.43)	5.00	1.00	8.00
Efficiency*	.85 (.64)	.77	.02	2.14
	Complex interpretation tasks (n = 52)			
Accuracy	.78 (.25)	.67	.00	1.00
Mental Load	6.31 (1.74)	7.00	1.00	9.00
Efficiency*	.77 (.64)	.58	.02	2.28

* Based on the Paas and Meriënboer (1992) equation.

Table 6

Model results RM-ANOVA

	df1, df2	*F*	*P*	N_G^2
		Dependant variable = Accuracy		
Compliance	1, 50	2.81	$.100$	$.025$
Complexity	1, 50	150.23	$< .001$	$.617$
Compliant:Complexity	1, 50	4.15	$.047$	$.043$
		Dependant variable = Mental effort		
Compliance	1, 50	1.18	$.283$	$.016$
Complexity	1, 50	30.26	$< .001$	.153
Compliant:Complexity	1, 50	0.14	0.706	$.706$
		Dependant variable = Efficiency		
Compliance	1, 50	0.13	.720	<.001
Complexity	1, 50	0.08	.785	<.001
Compliant:Complexity	1, 50	0.17	.680	.002

3.2 Appendix C – the survey

INFORMED CONSENT

"**Over the rainbow: reviewing current inclusive figure design practices within the field of Psychology and investigating the instructional efficiency of WCAG 2.0 compliant figures using cognitive load theory**"

EC Code: PED-2122-S-0077

- I have read the information about the research/study. I have had ample opportunity to ask questions about it.
- I understand what the study is about, what is being asked of me, which consequences participation can have, how my data will be handled, and what my rights as a participant are.
- I understand that participation in the study is voluntary. I choose to participate. I can stop participating at any moment. If I withdraw from the study, I do not need to explain why. Withdrawing from the study will have no negative consequences for me.
- Below I indicate what I am consenting to.

Question 1
Consent to participate in the research:
[ ] Yes, I consent to participate; this consent is valid until 31-12-2022
[ ] No, I do not consent to participate

Question 2
Consent to processing my personal data:
[ ] Yes, I consent to the processing of my personal data as mentioned in the research information. I know that until 31-12-2022, I can ask to have my data withdrawn and erased. I can also ask for this if I decide to stop participating in the research.
[ ] No, I do not consent to the processing of my personal data.

The researcher declares that the participant has received extensive information about the research.

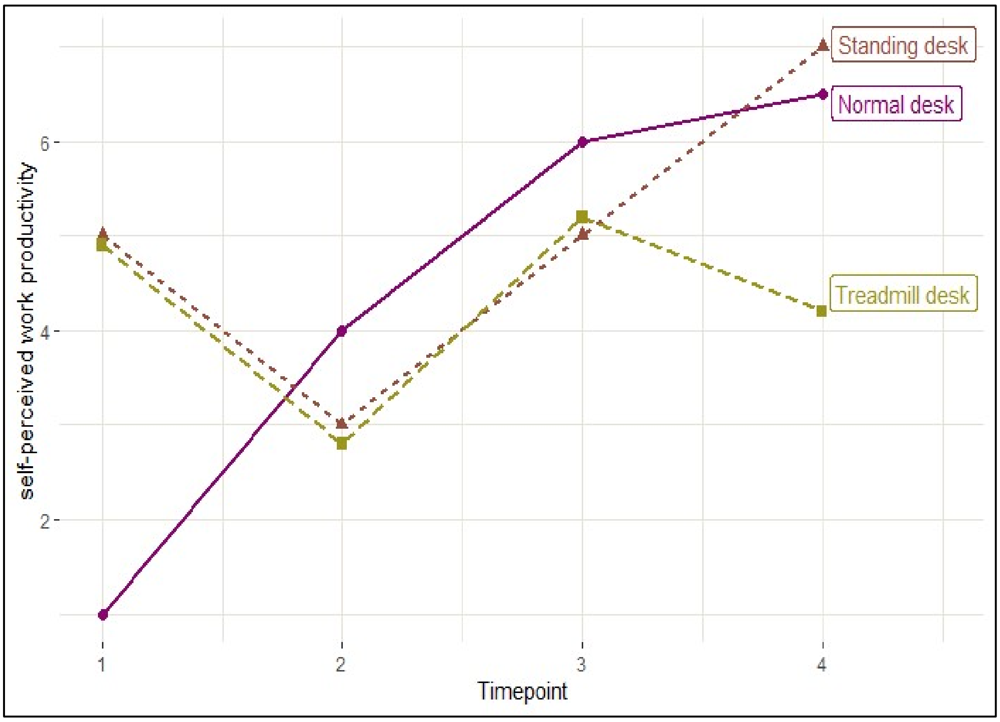
Question introduction – Groups A & B

The following two questions are based on a fictional study that studies the self-perceived productivity levels of employees based on their desks. The group of participants are divided into three possible types of desks: (1) a normal desk, (2) a Standing desk, and (3) a Treadmill desk.

Over four measurement points each three months apart, the participant was asked to rate their self-perceived productivity level ranging from 0 to 8, with 0 being no productivity change and 8 being a tremendous change in productivity.

Question 3 – Group A (Compliant)

Figure 8: Results from an experiment using three types of desks and how those types impact self-perceived productivity (vertical axis). With each time point spanning over three months (horizontal axis).



a) What is the average value of the treadmill desk group at timepoint 3?

b) What is the range of average values for the standing desk group?

- Between [participant answer] and [participant answer]

c) Which group shows a decreasing trend between timepoints 1 and 2, and increases in subsequent timepoints?

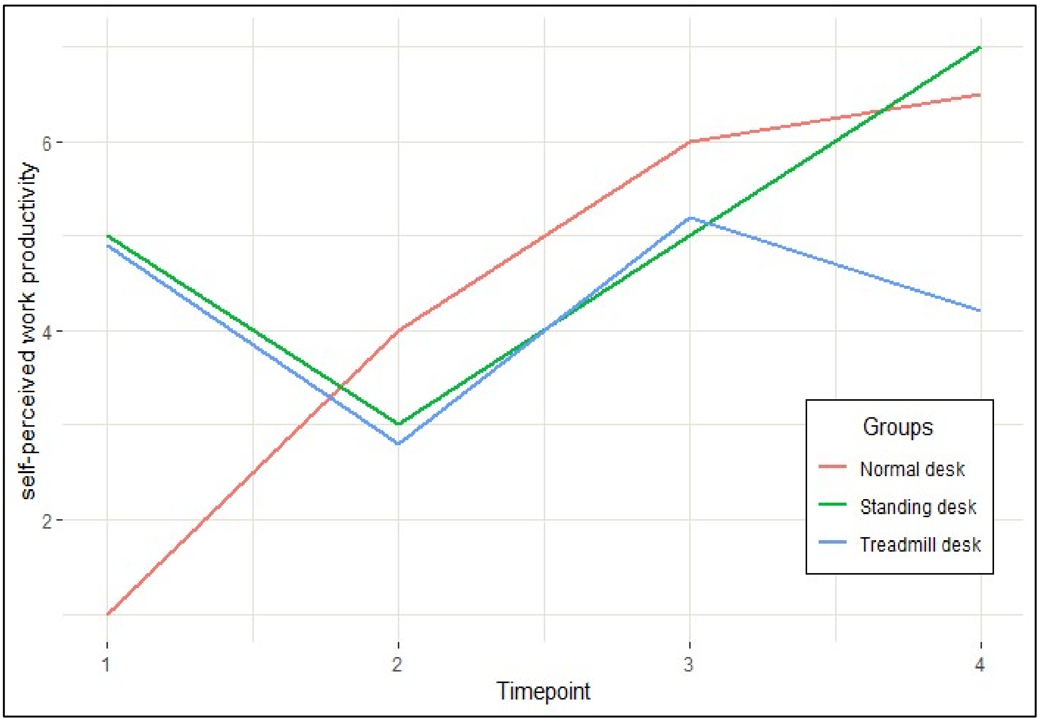
- Standing desk
- Normal desk
- Treadmill desk

d) How much mental effort did it cost to solve the three exercises?

- 1 = very, very low mental effort;
- 2 = very low mental effort;
- 3 = low mental effort;
- 4 = rather low mental effort;
- 5 = neither low nor high mental effort;
- 6 = rather high mental effort;
- 7 = high mental effort;
- 8 = very high mental effort;
- 9 = very, very high mental effort

Question 3 – Group B (Non-compliant)

Figure 9: Results from an experiment using three types of desks and how those types impact self-perceived productivity (vertical axis). With each time point spanning over three months (horizontal axis).



a) What is the average value of the treadmill desk group at timepoint 3?

b) What is the range of average values for the standing desk group?

- Between [participant answer] and [participant answer]

c) Which group shows a decreasing trend between timepoints 1 and 2, and increases in subsequent timepoints?

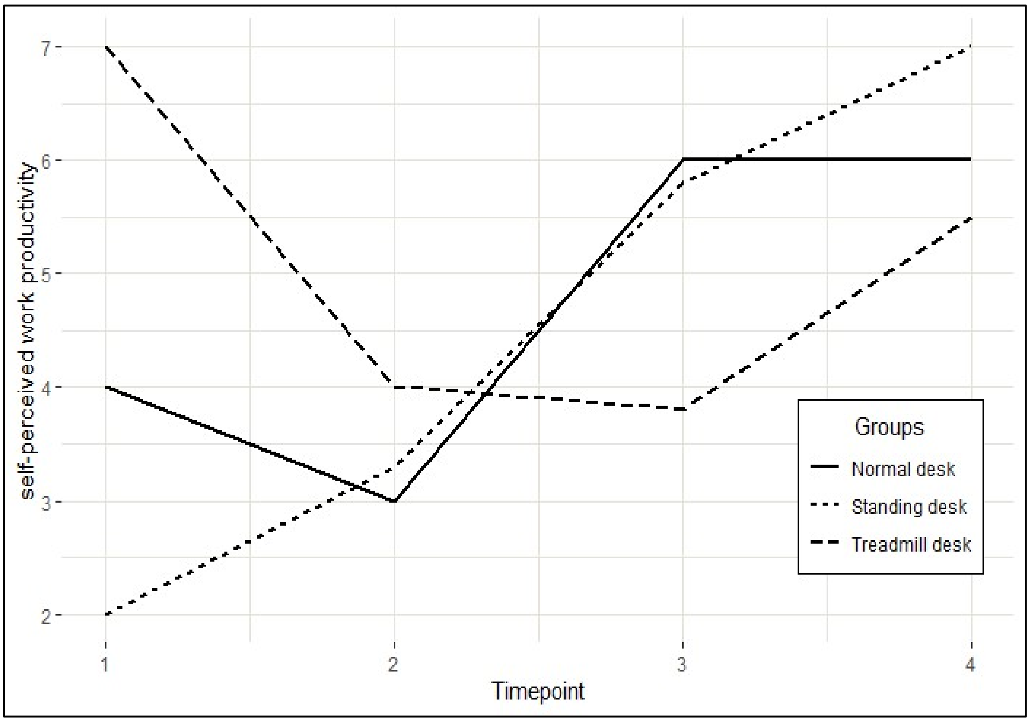
- Standing desk
- Normal desk
- Treadmill desk

d) How much mental effort did it cost to solve the three exercises?

- 1 = very, very low mental effort;
- 2 = very low mental effort;
- 3 = low mental effort;
- 4 = rather low mental effort;
- 5 = neither low nor high mental effort;
- 6 = rather high mental effort;
- 7 = high mental effort;
- 8 = very high mental effort;
- 9 = very, very high mental effort

Question 4 – Group A + B (greyscale compliant by default)

Figure 2: Results from an experiment using three types of desks and how those types impact self-perceived productivity (vertical axis). With each time point spanning over three months (horizontal axis).



a) What is the average value of the standing desk group at timepoint 3?

b) What is the range of average values for the normal desk group?

- Between [participant answer] and [participant answer]

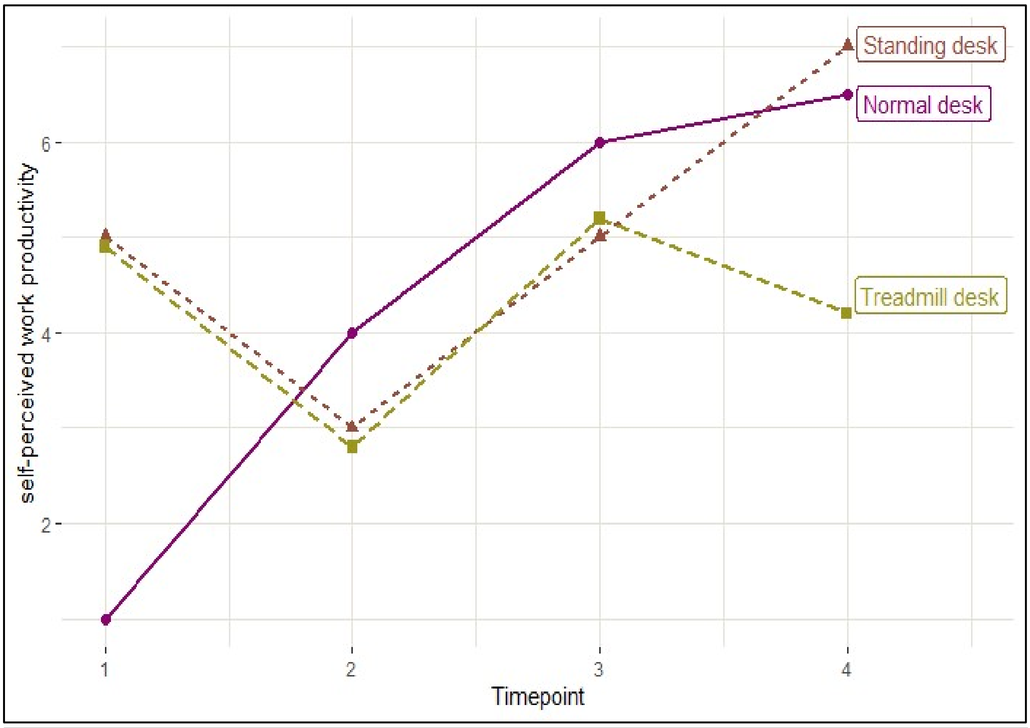
c) Which group shows a decreasing trend between timepoints 1 and 2, and increases in subsequent timepoints?

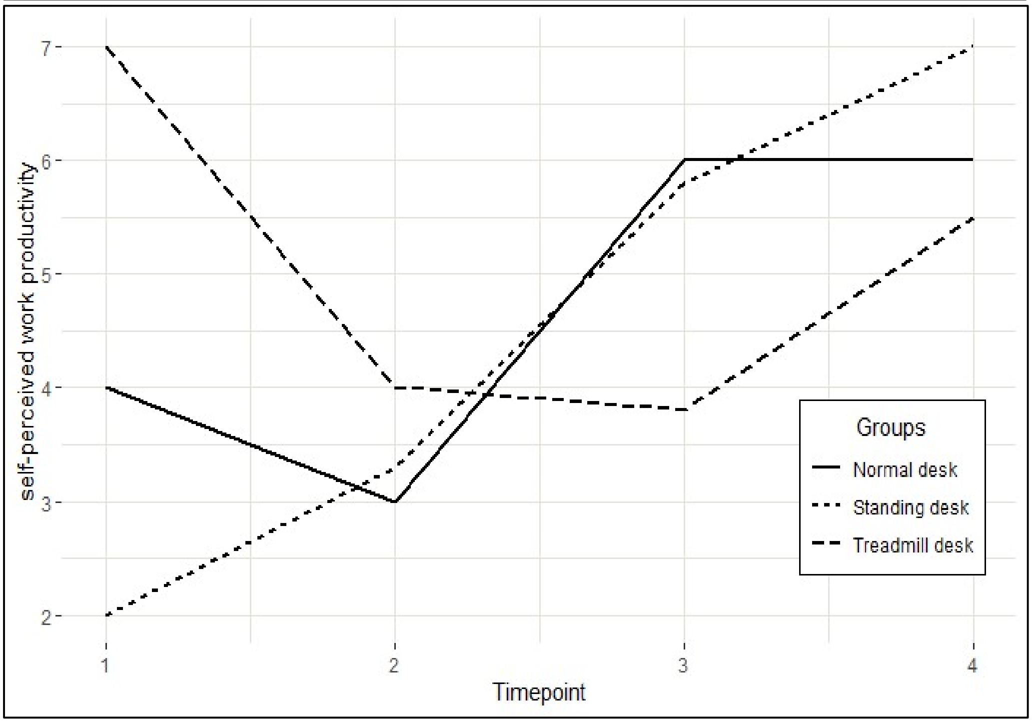
- Standing desk
- Normal desk
- Treadmill desk

d) How much mental effort did it cost to solve the three exercises?

- 1 = very, very low mental effort;
- 2 = very low mental effort;
- 3 = low mental effort;
- 4 = rather low mental effort;
- 5 = neither low nor high mental effort;
- 6 = rather high mental effort;
- 7 = high mental effort;
- 8 = very high mental effort;
- 9 = very, very high mental effort

Question 5 – Group A (Compliant)

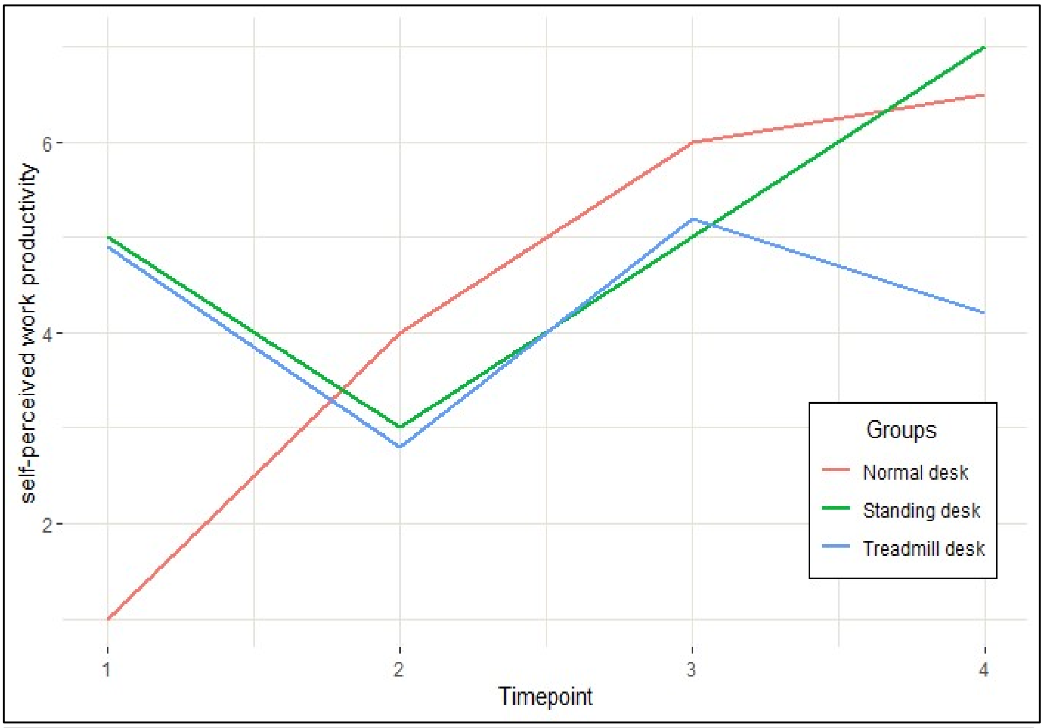


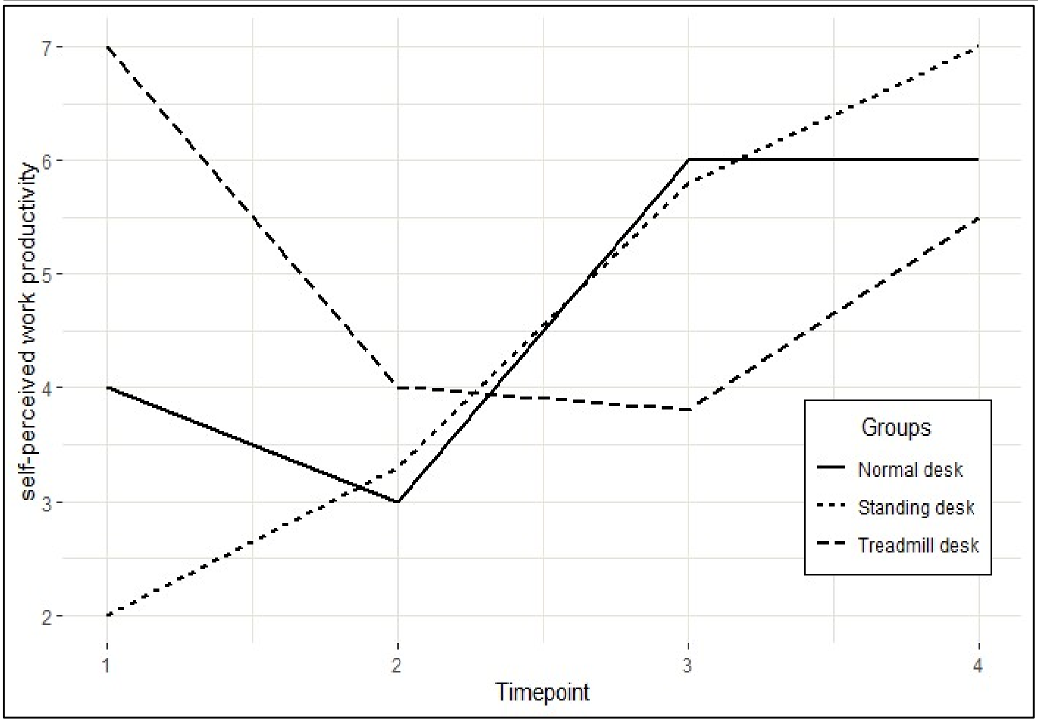


Which of the previously displayed data visualisations has your preference?

- Coloured variant, because [textbox]
- Grayscale variant, because [textbox]
- No preference, because [textbox]

Question 5 – Group A (Non-compliant)





Which of the previously displayed data visualisations has your preference?

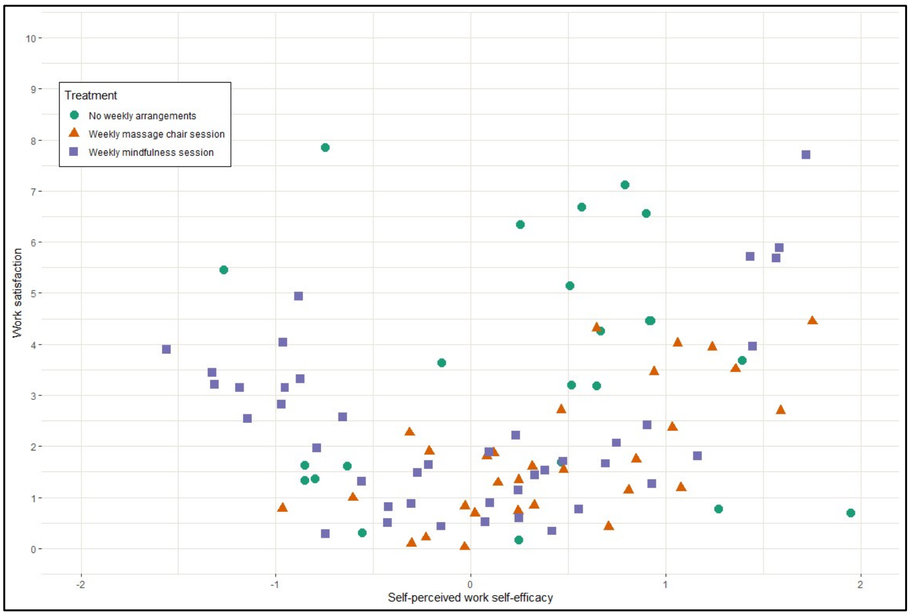
- Coloured variant, because [textbox]
- Grayscale variant, because [textbox]
- No preference, because [textbox]

Question introduction – Groups A & B
The following two questions are based on a fictional study which studies the self-perceived work efficacy based on different weekly working arrangements. The groups of participants are divided into (1) no weekly working arrangement, (2) weekly massage chair sessions, and (3) weekly mindfulness sessions.

Self-perceived work self-efficacy has scores on the horizontal axis ranging from -3 (no confidence in managing themselves in the workplace) to 3 (high confidence in managing themselves in the workplace. Work satisfaction is measured from -10 (extremely unsatisfied at work) and 10 (extremely satisfied at work).

Question 6 – Group A (Compliant)

Figure 10: *This data visualisation portrays participants' scores for three different working arrangements (no weekly arrangement, weekly mindfulness session, and weekly massage session) and how they impact the relation between self-perceived work self-efficacy (horizontal axis) and work satisfaction (vertical axis).*



a) What is the perceived trend for participants receiving mindfulness?

- Positive
- Neutral
- Negative

b) Make an educated guess about the correlation between work satisfaction and self-perceived work self-efficacy for the group weekly massage session.

- The correlation is between 0.0 and 0.25
- The correlation is between 0.25 and 0.50
- The correlation is between 0.50 and 0.75
- The correlation is between 0.75 and 1.00

c) What type of relationship exists between work satisfaction and self-perceived work self-efficacy for the group weekly mindfulness session?

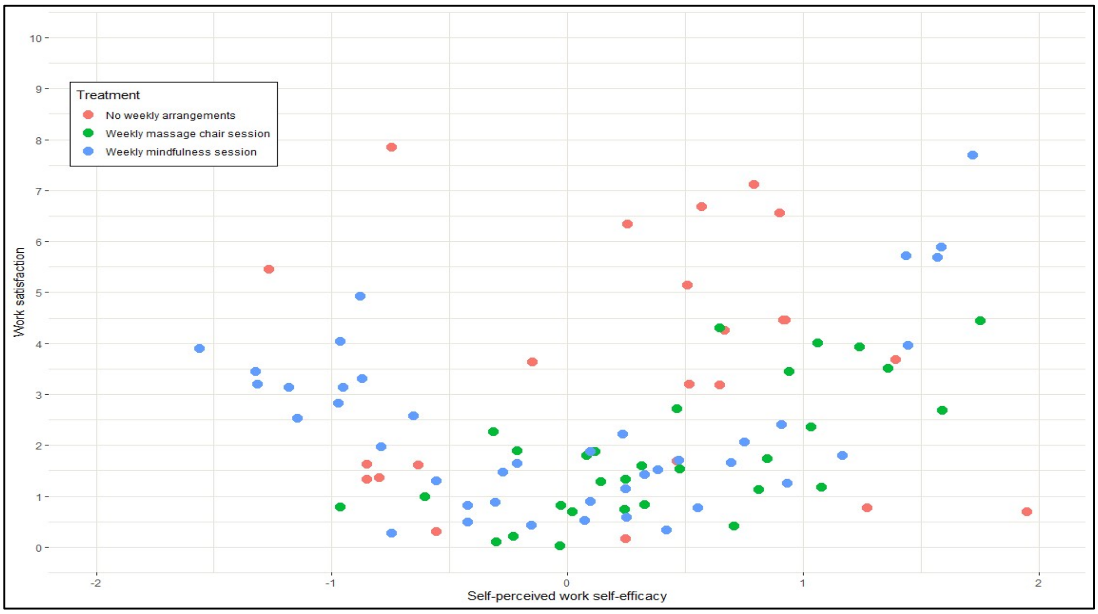
- a positive linear relationship
- a negative linear relationship
- non-linear relationship
- no relationship

d) How much mental effort did it cost to solve the three exercises?

- ○ 1 = very, very low mental effort;
- ○ 2 = very low mental effort;
- ○ 3 = low mental effort;
- ○ 4 = rather low mental effort;
- ○ 5 = neither low nor high mental effort;
- ○ 6 = rather high mental effort;
- ○ 7 = high mental effort;
- ○ 8 = very high mental effort;
- ○ 9 = very, very high mental effort

Question 6 – Group A (Non-compliant)

Figure 11: *This data visualisation portrays participants' scores for three different working arrangements (no weekly arrangement, weekly mindfulness session, and weekly massage session) and how they impact the relation between self-perceived work self-efficacy (horizontal axis) and work satisfaction (vertical axis).*



a) What is the perceived trend for participants receiving mindfulness?

- Positive
- Neutral
- Negative

b) Make an educated guess about the correlation between work satisfaction and self-perceived work self-efficacy for the group weekly massage session.

- The correlation is between 0.0 and 0.25
- The correlation is between 0.25 and 0.50
- The correlation is between 0.50 and 0.75
- The correlation is between 0.75 and 1.00

c) What type of relationship exists between work satisfaction and self-perceived work self-efficacy for the group weekly mindfulness session?

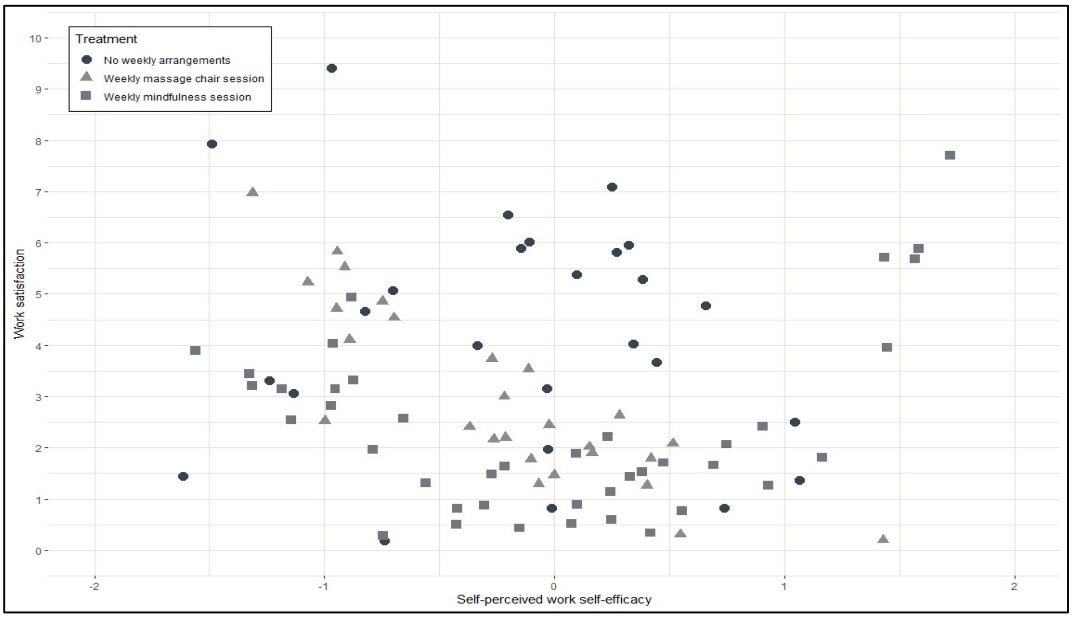
- a positive linear relationship
- a negative linear relationship
- non-linear relationship
- no relationship

d) How much mental effort did it cost to solve the three exercises?

- 1 = very, very low mental effort;
- 2 = very low mental effort;
- 3 = low mental effort;
- 4 = rather low mental effort;
- 5 = neither low nor high mental effort;
- 6 = rather high mental effort;
- 7 = high mental effort;
- 8 = very high mental effort;
- 9 = very, very high mental effort

Question 6 – Group A + B (greyscale compliant by default)

Figure 12: *This data visualisation portrays participants' scores for three different working arrangements (no weekly arrangement, weekly mindfulness session, and weekly massage session) and how they impact the relation between self-perceived work self-efficacy (horizontal axis) and work satisfaction (vertical axis).*



a) What is the perceived trend for participants receiving weekly massage sessions?

- Positive
- Neutral
- Negative

b) Make an educated guess about the correlation between work satisfaction and self-perceived work self-efficacy for the group weekly mindfulness session.

- The correlation is between 0.0 and 0.25
- The correlation is between 0.25 and 0.50
- The correlation is between 0.50 and 0.75
- The correlation is between 0.75 and 1.00
-

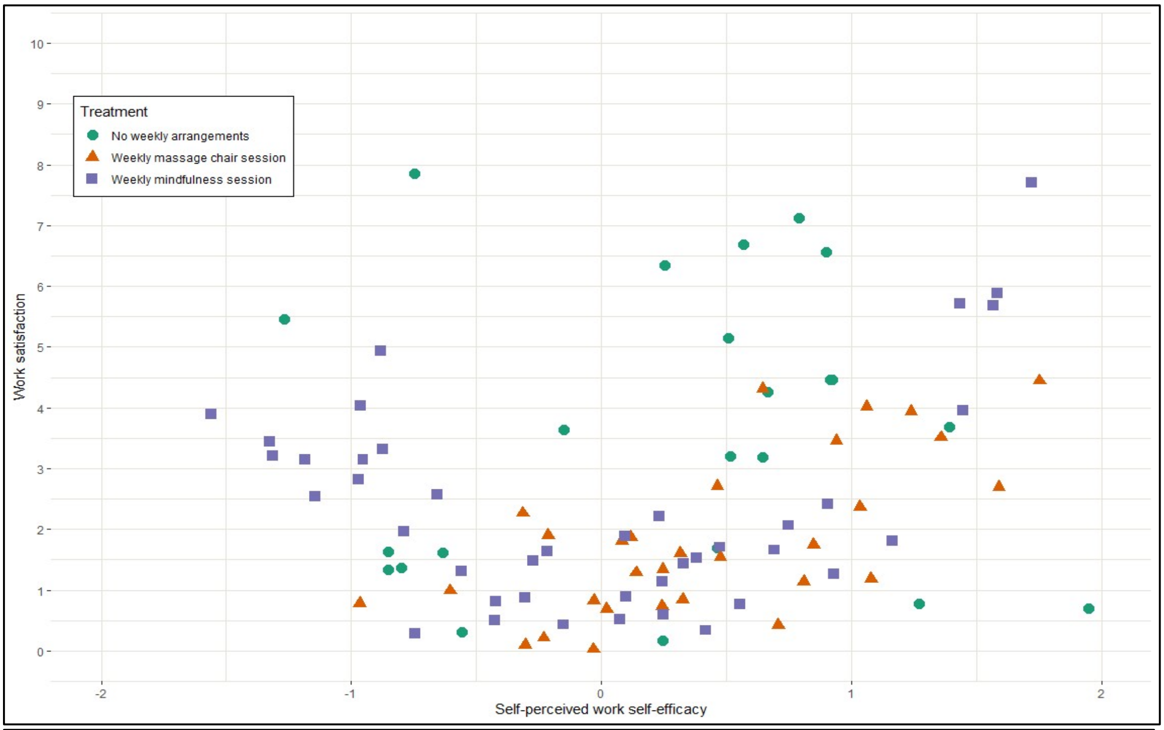
c) What type of relationship exists between work satisfaction and self-perceived work self-efficacy for the group weekly massage session?

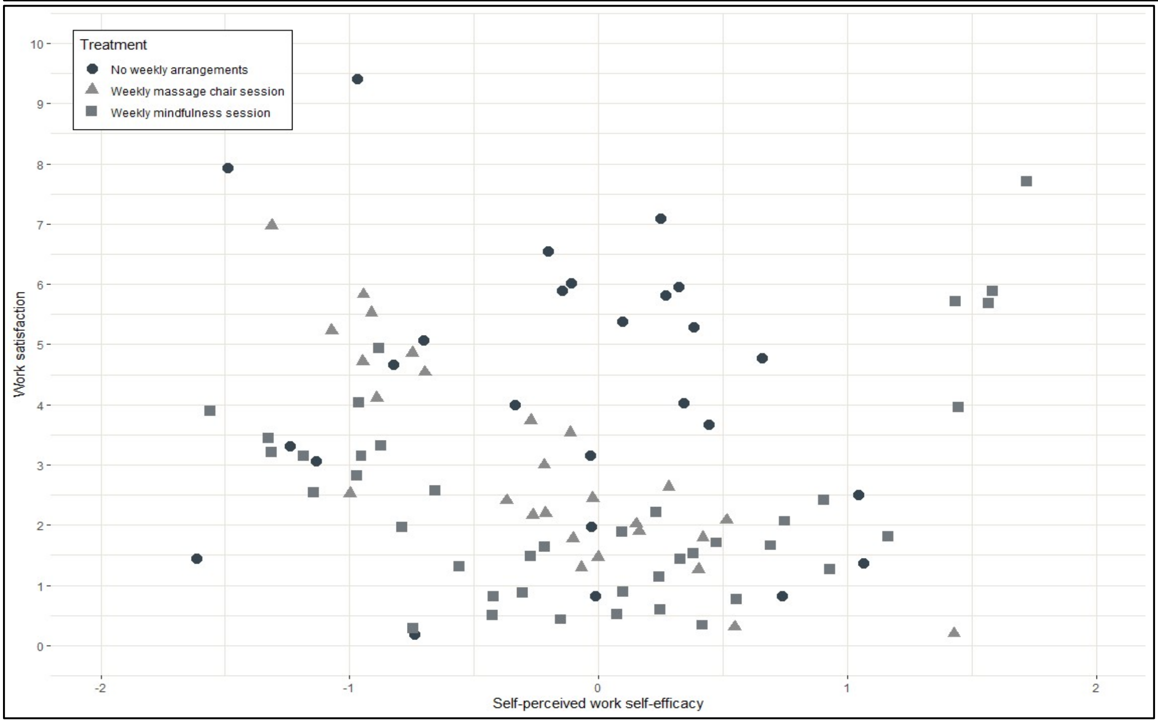
- a positive linear relationship
- a negative linear relationship
- non-linear relationship
- no relationship

d) How much mental effort did it cost to solve the three exercises?

- 1 = very, very low mental effort;
- 2 = very low mental effort;
- 3 = low mental effort;
- 4 = rather low mental effort;
- 5 = neither low nor high mental effort;
- 6 = rather high mental effort;
- 7 = high mental effort;
- 8 = very high mental effort;
- 9 = very, very high mental effort

Question 7 – Group A (Compliant)

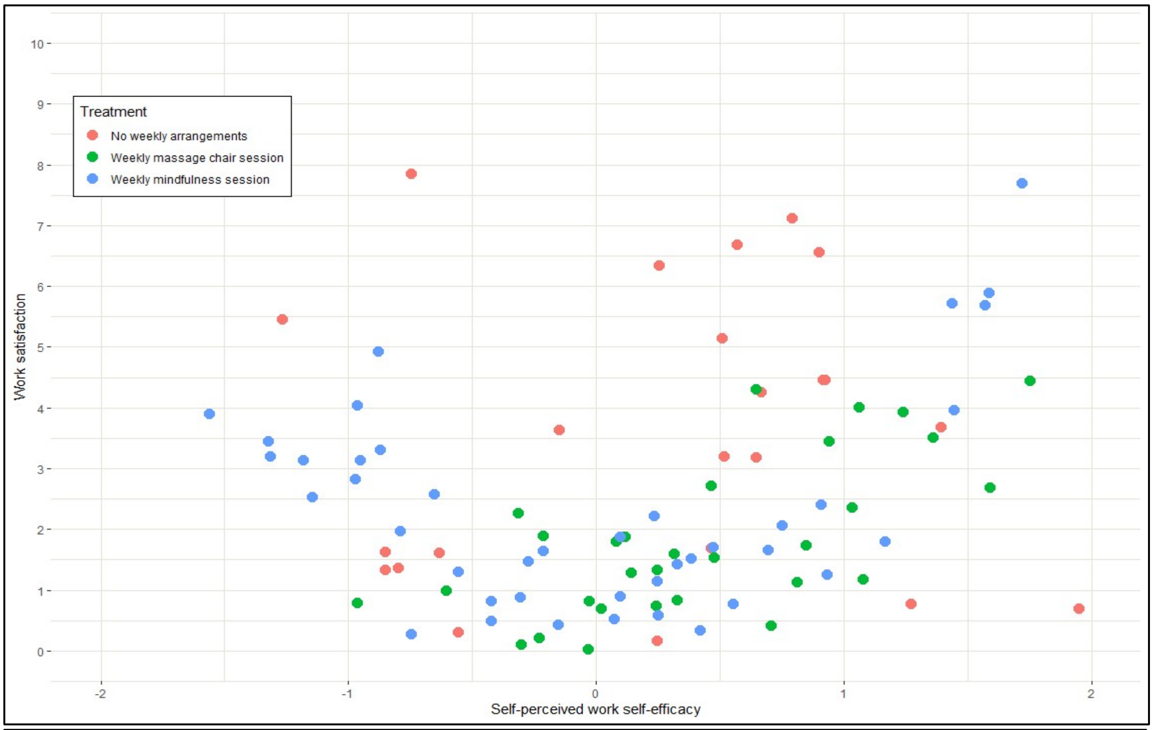


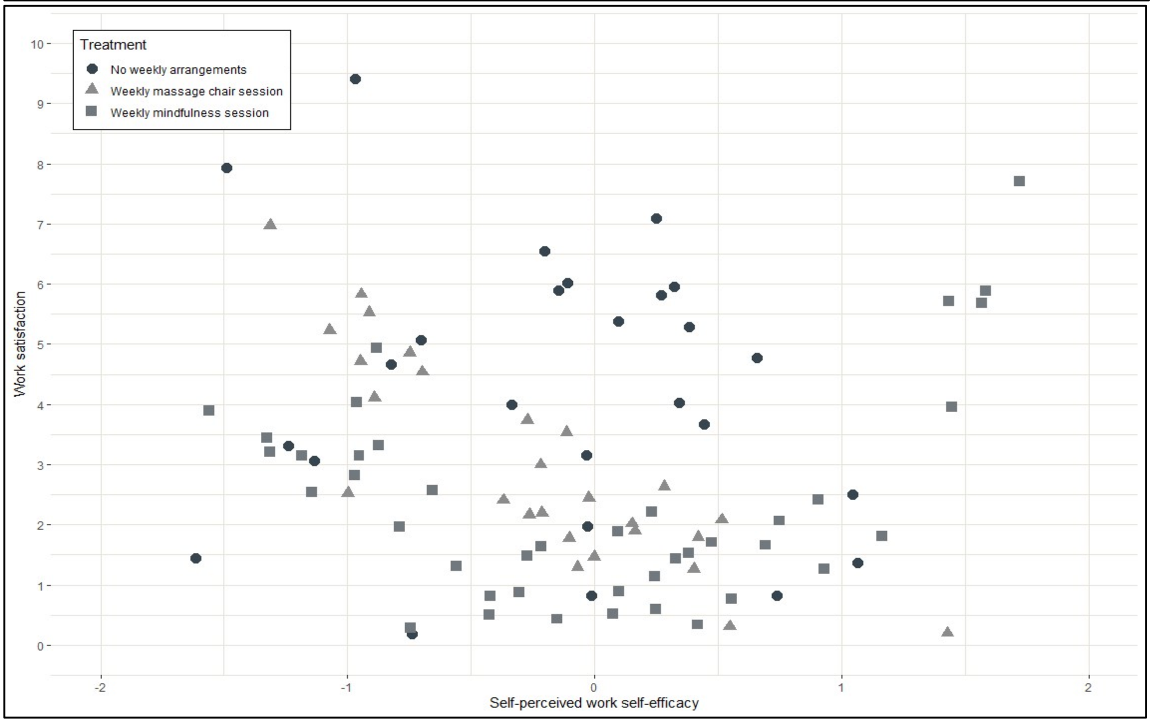


Which of the previously displayed data visualisations has your preference?

- Coloured variant, because [textbox]
- Grayscale variant, because [textbox]
- No preference, because [textbox]

Question 7 – Group B (Non-Compliant)





Which of the previously displayed data visualisations has your preference?

- Coloured variant, because [textbox]
- Grayscale variant, because [textbox]
- No preference, because [textbox]

Question 8 – Group A + B
Do you have any comments, please leave them here:

- [Participant answer]