Title

Physiological Processing, Perceived Effort, and Recall Performance for Information from Social Media Scrolling Feeds

Author

Loh, Zoe

Publication Date

2023

UNIVERSITY OF CALIFORNIA, MERCED

Physiological Processing, Perceived Effort, and Recall Performance for Information from Social Media Scrolling Feeds

A Thesis submitted in partial satisfaction of the requirements for the degree of Master of Science

in

Management of Complex Systems

by

Zoe Loh

Committee in charge:

Professor Spencer Castro, Chair
Professor Paul Maglio
Professor Lace Padilla

2023

The Thesis of Zoe Loh is approved and it is acceptable in quality and form for publication on

microfilm and electronically:

Professor Paul Maglio

Professor Lace Padilla

Professor Spencer Castro, Chair

University of California, Merced

2023

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank my advisor, Dr. Spencer Castro for his support and guidance on this work. Thank you to the members of the TECH lab. Without them, this work would not be possible. I am also grateful to my committee members, Dr. Lace Padilla and Dr. Paul Maglio for their valuable feedback and support. Thank you to my friends and family for their understanding and support throughout my graduate school journey.

Abstract

Physiological Processing, Perceived Effort, and Recall Performance for Information from Social Media Scrolling Feeds

A Thesis submitted in partial satisfaction of the requirements for the degree of Master of Science in Management of Complex Systems, University of California, Merced, 2023

Professor Spencer Castro, Committee Chair

As social media scrolling feeds become a major source of information for the American public, understanding how individuals process this information will become essential. The present study assesses how the format of different visual formats and media impact how people process and remember this information. Across two studies, we recorded participant eye movements and pupil sizes while reading sections of a climate change report in the form of 1) a digital PDF and 2) a simulated social media scrolling feed. This report was presented on a computer screen or as a physical paper. At the end of each block, participants answered multiple-choice questions and completed the NASA Task Load index as a measure of perceived effort. We found that participants reported greater mental demand for the PDF compared to the scrolling feed and for the digital version compared to the physical version. However, we could not conclude a difference in accuracy between the two formats or the two media.

Introduction

As social media scrolling feeds become a major source of information for the American public, understanding how individuals process (take in, store, and select a response to) this information will become essential. A survey conducted in 2020 revealed that around 53% of American adults at least sometimes use social media as a news source (Shearer & Mitchell, 2021). How people make decisions and draw conclusions from social media is subject to psychological principles such as memory and information processing limits (Rodriguez et al., 2014). These conclusions and decisions then in turn impact everyday behaviors such as consumer behaviors (Voramontri & Klieb, 2019).

Background

We first establish the limited capacity of human cognitive resources, such as human memory capacity and processing ability, because processing and recalling information requires these limited resources (e.g. Kahneman, 1973; Miller, 1956). The next section introduces the concept of cognitive workload and the factors that can affect the level of workload for a task. The following section provides more detail on different types of cognitive workload measures and why multiple, converging measures are needed. Finally, the last sections cover how the format and medium in which information is presented may affect recall and processing.

Cognitive Resource Limits. There are limits to human memory capacity (the amount of information that can be held in memory at a given time) and processing ability (ability to take in, store, and select a response to information)(e.g. Kahneman, 1973; Miller, 1956). *Working memory*, a temporary store of information that allows for consolidation and processing of more complex, higher-level semantic information, is limited by several factors (Chai et al., 2018). According to the *embedded-processes model*, working memory is comprised of different cognitive processes that allow information from long-term memory storage and short-term memory storage (an activated subset of long-term memory storage) to remain accessible for performing mental tasks (Cowan, 1999). The cognitive processes involved in working memory, the activation of information in memory, and the focus of attention are constrained by time limits and capacity limits respectively (Cowan, 1999). One of the methods for expanding the amount of information that can be held in memory at one time with these limitations is chunking, *Chunking* is the process of grouping pieces of information being held in working memory that allows greater amounts of information to be stored in each unit stored in memory (Simon, 1974; Thalmann et al., 2019). For example, social media such as Twitter (also known as X) restricts the characters and therefore the information density to 280 characters per post (*Counting Characters*, 2023). In addition, social media posts bound the information within a uniform color (following the Gestalt principle that items are perceptually grouped using similar color) and a bounded line/Frame (Holmes et al., 2018; Peterson & Berryhill, 2013) While memory capacity and information processing are limited by various constraints, there are strategies such as chunking to increase what we hold in memory storage within those limitations.

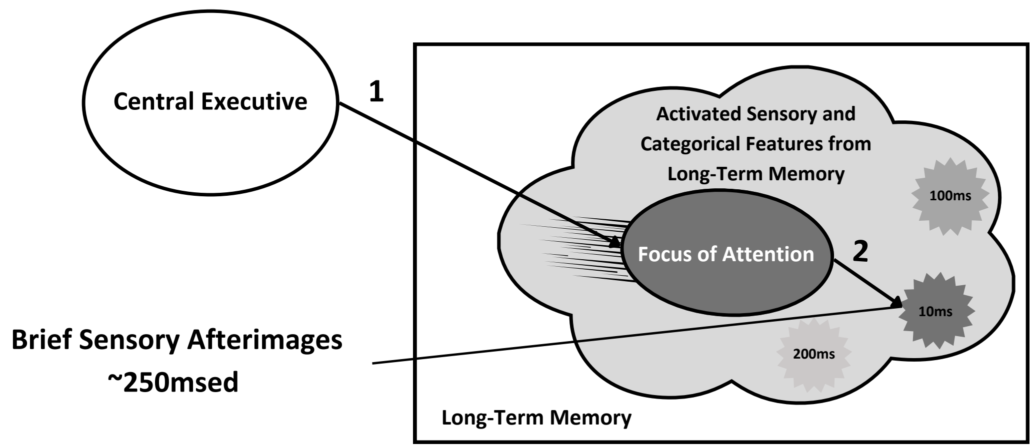

Figure 1. Adapted from Embedded-Processes Model (Cowan, 1999)

Cognitive Workload. *Cognitive workload* is the mental effort that is required to complete a task (Thomson & Oppenheimer, 2022). Greater demand on the limited attentional resources available for information processing results in higher workload (Barrouillet et al., 2007). Additional cognitive workload can be induced by dividing attentional resources between a primary and secondary task such as counting backwards by 3s while driving a car (Castro et al., 2019). This results in slowing the rate of *evidence accumulation*, which is the process of gathering information towards making a decision until a threshold is reached and a response is produced; higher workload also increases the amount of evidence required to reach the threshold (Castro et al., 2019).

Measurement. Researchers have used a variety of self-report (subjective), physiological, and performance-based measures of cognitive workload. One common subjective measure of cognitive workload is the NASA Task Load Index (NASA TLX), a series of Likert scale questions used to assess perceived cognitive workload for a given task (Hart & Staveland, 1988). Physiological measures of cognitive workload include electroencephalography (EEG), neuroimaging, and pupillometry (Chen & Epps, 2014; Gevins & Smith, 2003; Just et al., 2003). Performance-based measures evaluate cognitive workload as a function of behavioral; cognitive workload increases as performance decreases (e.g. Bufano et al., 2022; Castro et al., 2019).

Much of the previous work relies on using a single approach in measuring cognitive workload. However, there are advantages to using multiple measures. While multiple workload measures have converged for some studies (e.g. Kramer, 1986), other work has revealed that measures can diverge from one another even with within-measure reliability (Matthews et al., 2015). Indeed, one approach alone does not fully capture workload and various approaches may assess different latent constructs underlying workload (Matthews et al., 2020). Therefore, it is necessary to have multiple measures to investigate various aspects of workload and reveal new information.

In the present study, we will use the NASA TLX as a subjective measure of workload and incorporate pupil size as a physiological indicator of workload. As cognitive workload increases, pupil size also increases (Biondi et al., 2023).Using these measures, we can compare the cognitive workload induced by presenting information in different formats and media.

Format. While previous research has compared portable document formats (PDFs) to webpages, less is known about how they compare to more modern digital formats of information such as social media feeds (Nielsen Norman Group, 2020). As social media tends to summarize and chunk (group) pieces of information into posts, we hypothesize that this may improve efficiency of recall. Format choices have the potential to impact the grouping of information. For instance, having spatially segmented pieces information (frames) can benefit memory for a spatial layout (Holmes et al., 2018). As the information contained within a social media post is separated from the rest of the page with a frame, this format may facilitate recall. In addition, Gestalt principles of similarity and proximity state that objects with similar characteristics (e.g. color, shape) and are closer together spatially will be grouped together perceptually, which may benefit working memory (Peterson & Berryhill, 2013). As social media posts have information bound in the same color and the text within the post spaced closely together, the social media format may provide an advantage in terms of recall efficiency.

Medium. In addition to format, medium may also impact recall and processing of information. According to *cognitive fit theory*, problem solving depends on the relationship between a given problem-solving task and the way the task is represented externally (Vessey & Galletta, 1991). The perceived fit (alignment) between technology and task can therefore influence performance (Vessey & Galletta, 1991). As PDFs are created to be read on printed paper and social media feeds are created for digital displays, we predict that recall performance will be better for PDFs than social media feeds when presented in a physical medium and vice versa for a digital medium.

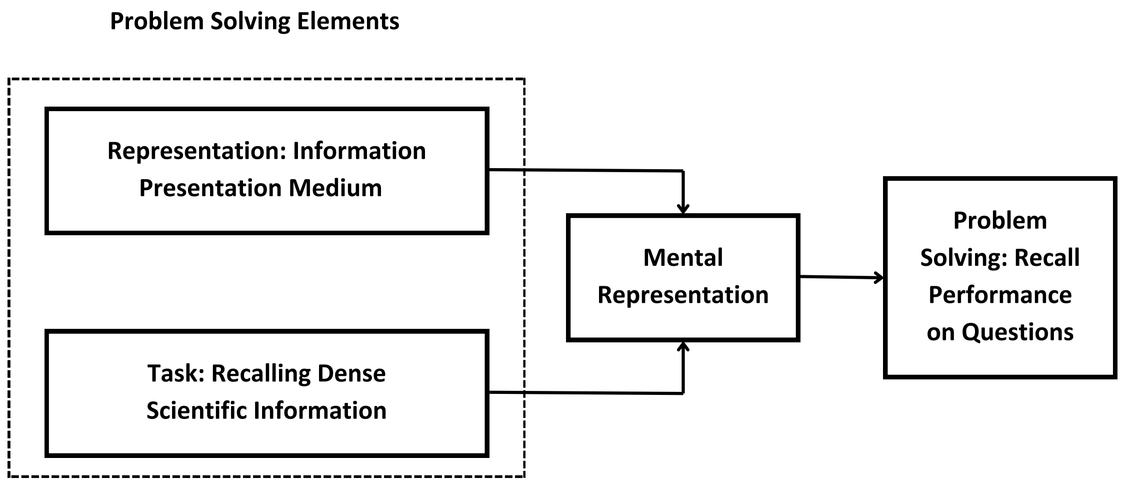

Figure 2. Adapted from Cognitive Fit Theory (Vessey & Galletta, 1991).

Present Study

The present study assesses how the format of different visual formats and media impact how people process and remember this information.

Our hypotheses are:

- H1: Recall accuracy will be higher for the scrolling feed format compared to the PDF format.
- H2: Perceived processing effort will be lower for the scrolling feed format compared to the PDF format.
- H3: Physiological processing effort will be lower for the scrolling feed format compared to the PDF format.

Pilot Study

Method

Participants. Participants were recruited through the online participant platform Prolific and received monetary compensation at the rate of $15 per hour for their participation in the study (Prolific, 2023). 100 participants (40 male, 59 female, 1 other) completed the study. Participant age averaged 25.91 years (*SD* = 13.27 years).

Stimuli. In order to investigate information recall and processing for a relevant news topic, this study used the Intergovernmental Panel on Climate Change (IPCC)'s 6th climate change report summary as the content of the simulated social media feed and the traditional information format (PDF file) (IPCC, 2021). The scrolling feed had a

“Twitter-like” appearance with posts following one another vertically down the page since Twitter is among one of the most popular text-based social media platforms among young adults in the United States (Auxier & Anderson, 2021). The posts were condensed summaries of subsections from the original IPCC summary report, each limited to 280 characters to be consistent with Twitter’s character limit restrictions (*Counting Characters*, 2023). To control for the vocabulary level of the text, the wording was not changed when summarizing the original report for generating the posts. Visualizations from the report were included in the scrolling feed as separate posts with condensed versions of the figure caption as text. The stimuli were presented using a Qualtrics survey that required a minimum screen width of 1200 pixels to ensure participants could read the material without scrolling horizontally (Qualtrics, 2023).

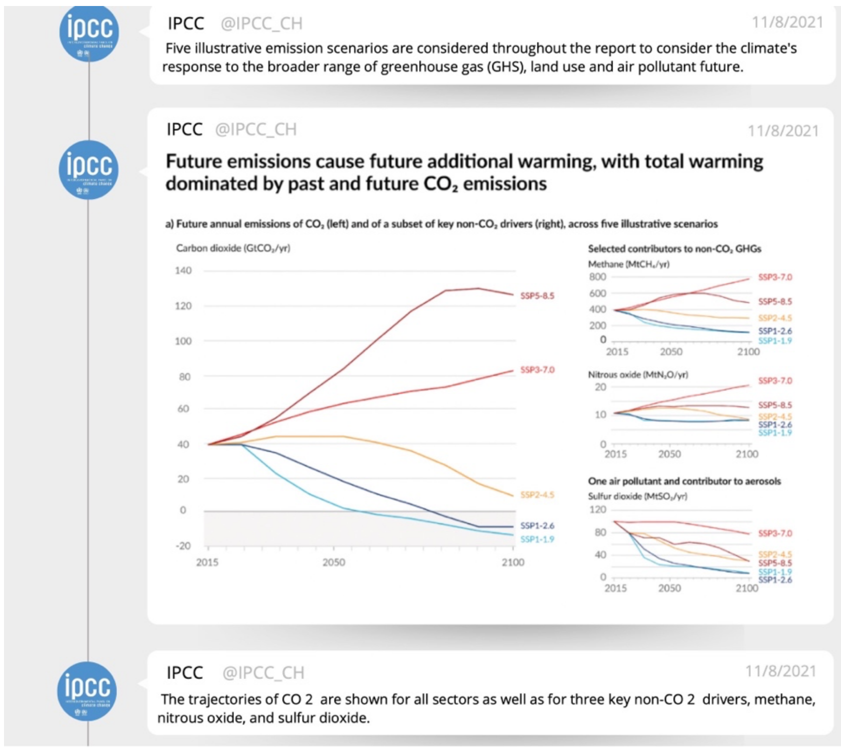

Figure 3. Example of the simulated social media feed format version of the IPCC report.

Procedure. After giving consent to participate in the study, each participant experienced both formats in a fully within-participants design. Participants read one section of the IPCC report in the form of a PDF file in one block and completed another block by reading a different section of the report as a simulated social media scrolling feed. Scrolling was self-paced for both formats. At the end of each block, participants answered multiple-choice questions about the information from that section and completed the NASA Task Load Index (TLX) (Hart & Staveland, 1988). Block order was counterbalanced between participants, as well as the content represented as a PDF or a scrolling feed.

Measures. Participant responses were collected for accuracy of 20 multiple-choice recall questions about the content of the report (the full set of questions can be found in Appendix A). 13 of the questions referred to text in the report and the remaining 7 questions referred to visualizations. The NASA TLX was taken after each block as a measure of perceived cognitive effort for the different presentations of information (Hart & Staveland, 1988).

Analysis. We utilized a binomial general linear mixed effects model to determine the relationship between information format and recall accuracy. We specified the fixed effect as format (i.e., PDF vs Scrolling Feed) and random intercepts for participant number (e.g., 1-100) and question number (e.g., 1-10). We also used a general linear mixed effects model to determine the relationship between information format and perceived effort (i.e., the NASA TLX). We specified the fixed effect as format and random intercepts for each participant.

Results

Accuracy. In our study, we found that participants responded more accurately to the multiple-choice recall questions for the scrolling feed format compared to the PDF format ($b = 0.23$, $SE = 0.10$, $z = 2.34$, $p = .02$, 95% CI [0.04, 0.43]). On average, participants were 1.26 times more accurate for the scrolling feed condition compared to the PDF condition.

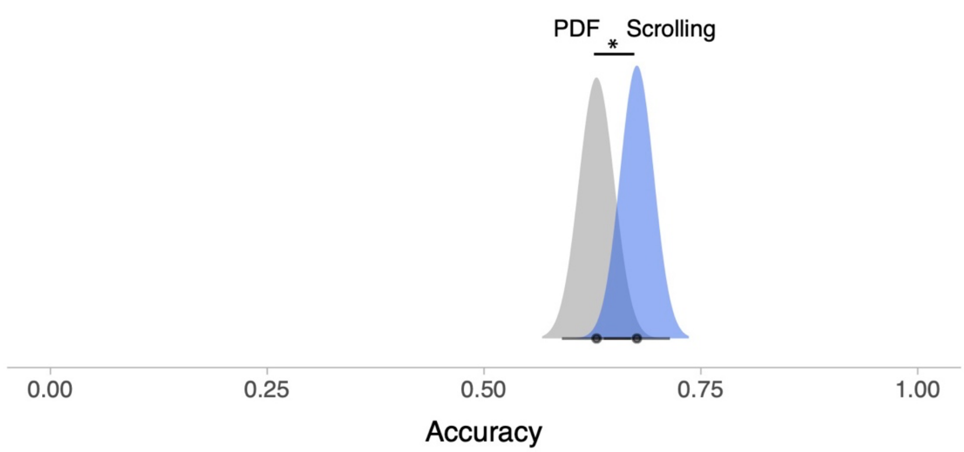

Figure 4. Distribution plot of recall accuracy by format. Accuracy is on the x-axis and density is on the y-axis. The plotted distributions are based on standard error. The black lines at the bottom of each curve represent the standard error and each curve represents if the standard error was 95% of the distribution. The black dots represent the means.

Perceived effort. Participants also reported lower mental demand for the scrolling feed compared to the PDF ($b = -0.77$, $SE = 0.22$, $t = -3.60$, $p < .001$, 95% CI [-

1.19, -0.35]). This effect was significant for three of the NASA Task Load Index Sub-Categories: Mental (How mentally demanding was the task?; $b = -1.23$, $SE = 0.24$, $t = -5.03$, $p < .001$, 95% CI [-1.71, -0.75]), Frustration (How discouraged or stressed did you feel during the task?; $b = -1.20$, $SE = 0.41$, $t = -2.93$, $p = .004$, 95% CI [-2.01, -0.40]), and Effort (How much effort did you put into completing the task?; $b = -1.12$, $SE = 0.35$, $t = -3.17$, $p = .002$, 95% CI [-1.81, -0.42]).

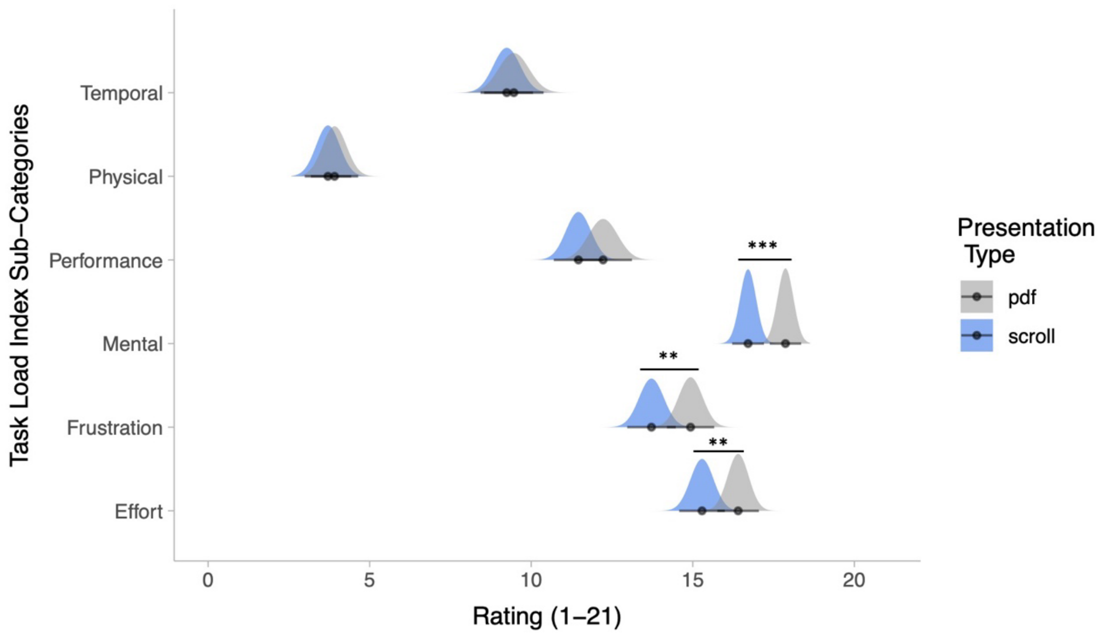

Figure 5. Distribution plot of perceived effort ratings by format, separated by NASA Task Load Sub-Categories. Rating is on the x-axis and sub-category is on the y-axis. The plotted distributions are based on standard error. The black lines at the bottom of each curve represent the standard error and each curve represents if the standard error was 95% of the distribution. The black dots represent the means.

Discussion

In this pilot study, we found that information presented in a scrolling feed format was recalled with higher accuracy (H1) and required lower perceived processing effort (H2) compared to information presented in a PDF format. Our results support our hypothesis that scrolling feeds may allow for lower perceived effort and better recall. We posit that one mechanism of this difference may be the chunking of information in a social media scrolling feed, which could provide an advantage in the efficiency of recall for information presented in this format.

Experiment 1

After conducting the online pilot study, we created an in-person study to compare the accuracy of recall, perceived processing effort, and physiological processing effort (i.e., pupil size) for information presented in the format of a simulated social media scrolling feed and a PDF. Based on the results of our pilot study, we hypothesized that participants would have higher accuracy (H1) and lower perceived effort (H2) for information presented in the social media feed format compared to the PDF format. We registered these hypotheses on the Open Science Framework (https://osf.io/6fu3x/). Self-reported and physiological measures of workload tend to converge for single tasks (Recarte et al., 2008). Therefore, we hypothesized that participants would exhibit a smaller dilation response (e.g., lower processing effort) for the scrolling feed format compared to the PDF format (H3).

Method

Participants. After a power analysis determined the number of participants needed, 72 students (14 male, 56 female, 2 non-binary/ third gender) were recruited via the university participant pool. The average age was 21 years (*SD* = 4.28 years).

Stimuli and apparatus. Participant pupil size and eye movements were recorded with an EyeLink Portable Duo, which sampled both eyes at a rate of 2000 Hz with a spatial resolution of 0.01 ° (SR Research, 2017). A chin rest was used to stabilize participant head position. A 60 cm diagonal monitor (1920 x 1080 pixels) displayed the stimuli 73 cm away from the participant. The presented stimuli were identical to the pilot study.

Procedure. The procedure was identical to the pilot study except that the participants had their eye movements and pupil size recorded as they completed the task.

Measures. Mean pupil size was recorded during the task as a measure of physiological cognitive effort. All other measures were identical to the pilot study.

Analysis. To examine characteristics of the multiple-choice recall questions, we averaged recall accuracy scores from the pilot study to obtain a mean difficulty rating for each question. We utilized a binomial general linear mixed effects model to determine the relationship between information format and recall accuracy. We specified the fixed effect as presentation format and random intercepts were specified for participants (e.g., 1-72) and question number (e.g., questions 1-10). We also implemented a general linear mixed effects model to determine if the information format predicted perceived effort. Our model specified the fixed effect as format and random intercepts for participants. We then specified a model to determine if information format and phase (e.g., study or test) predicted pupil size with participant as a random effect. Finally, we specified a model to determine the relationship between question characteristics (e.g., content and difficulty) and recall accuracy with random intercepts for participants.

Results

Accuracy. We could not conclude a difference in accuracy for multiple-choice recall question responses between the scrolling feed format and the PDF format (b = 0.13, SE = 0.12, z = 1.08, p =.28, 95% CI [-0.10, 0.36]).

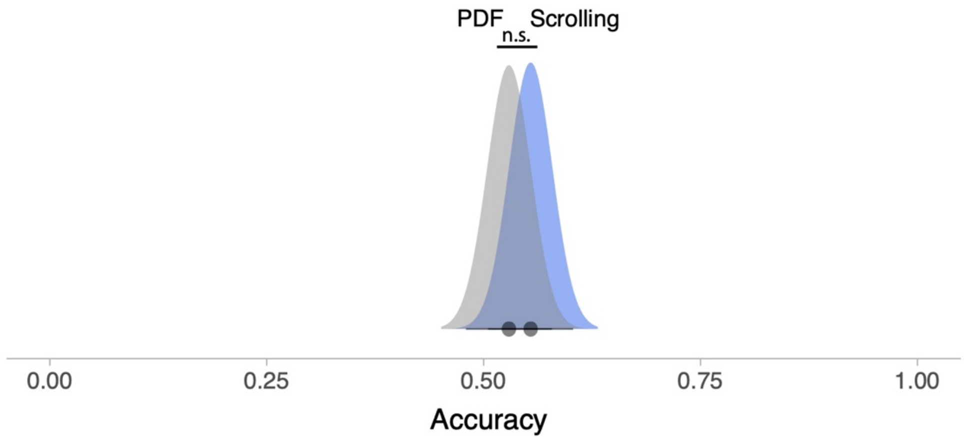

Figure 6. Distribution plot of recall accuracy by format. Accuracy is on the x-axis and density is on the y-axis. The plotted distributions are based on standard error. The black lines at the bottom of each curve represent the standard error and each curve represents if the standard error was 95% of the distribution. The black dots represent the means.

Perceived effort. Participants reported lower perceived effort for the scrolling feed compared to the PDF (b = -1.26, SE = 0.28, t = -4.55, p < .001, 95% CI [-1.80, -0.72]). This effect was significant for four of the NASA Task Load Index Sub-Categories: Physical (How physically demanding was the task?; b = -1.16, SE = 0.41, t = -2.83, p = .006, 95% CI [-1.97, -0.35], 95% CI [-2.93, -1.17]), Mental (How mentally demanding was the task?; b = -2.05, SE = 0.45, t = -4.58, p < .001, 95% CI [-2.93, -1.17]), Frustration (How discouraged or stressed did you feel during the task?; b = -2.26, SE = 0.56, t = -4.04, p < .001, 95% CI [-3.37, -1.16]), and Effort (How much effort did you put into completing the task?; b = -1.31, SE = 0.56, t = -2.35, p =.02, 95% CI [-2.42, -0.21]).

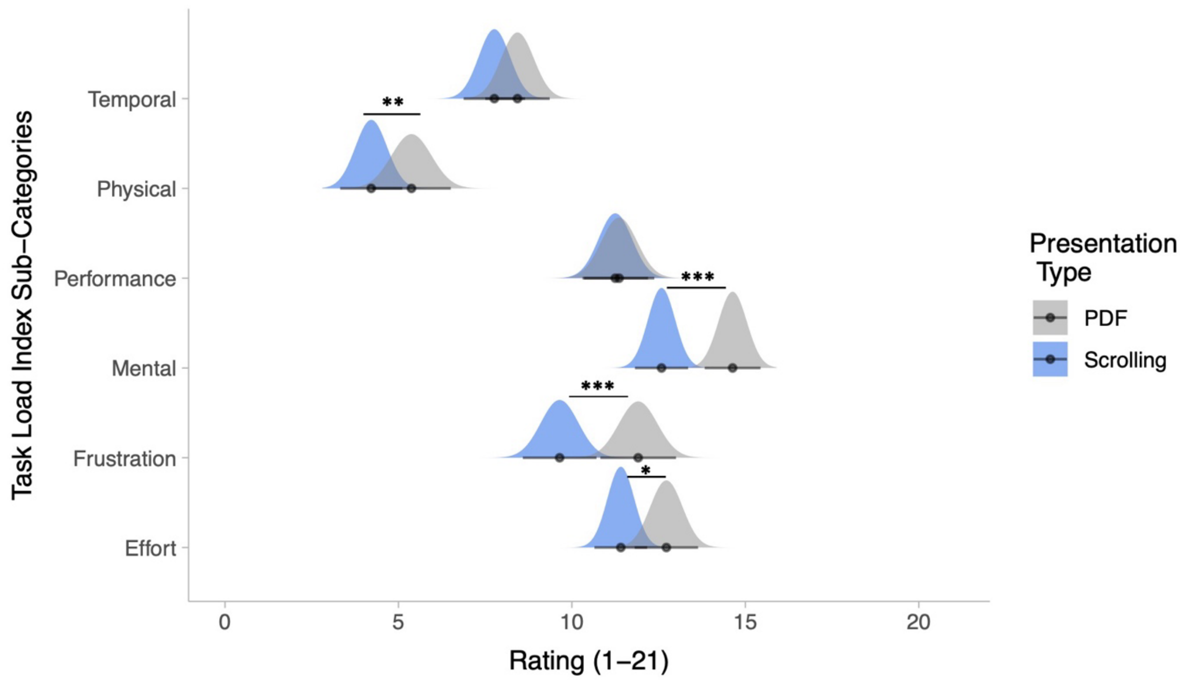

Figure 7. Distribution plot of perceived effort ratings by format, separated by NASA Task Load Sub-Categories. Rating is on the x-axis and sub-category is on the y-axis. The plotted distributions are based on standard error. The black lines at the bottom of each curve represent the standard error and each curve represents if the standard error was 95% of the distribution. The black dots represent the means.

Physiological effort. Participants had larger mean pupil size (i.e., greater cognitive effort) for the scrolling feed compared to the PDF (b = 16.33, SE = 3.73, t = 4.38, p < .001, 95% CI [9.02, 23.63]). Participants also had larger mean pupil size for the test phase compared to the study phase (b = 10.31, SE = 3.73, t = 2.77, p = .006, 95% CI [3.01, 17.63]).

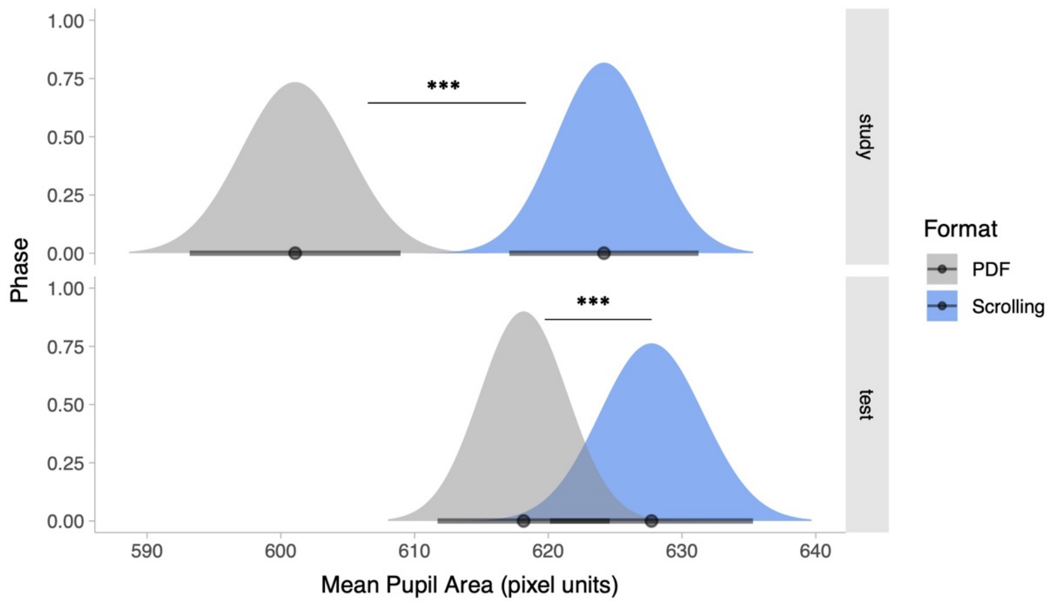

Figure 8. Distribution plot of mean pupil size by format, separated by phase. Mean pupil size is on the x-axis and phase is on the y-axis. The plotted distributions are based on standard error. The black lines at the bottom of each curve represent the standard error and each curve represents if the standard error was 95% of the distribution. The black dots represent the means.

Exploratory Analyses. Average question accuracy was 0.54 (*SD* = 0.20). Question difficulty was accounted for in the model by including the average accuracy from the pilot study for each question as a baseline. We found a positive relationship between baseline accuracy and recall accuracy (b = 5.03, *SE* = 0.13, z = -3.78, p < .001, 95% CI [3.96, 6.14]). After accounting for question difficulty, we did not find a statistically significant effect of question content on recall accuracy (b = -1.05, *SE* = 0.62, z = -1.70, p = .09, 95% CI [-2.28, 0.14]). We also did not find a statistically significant interaction between question content and question difficulty (b = 1.74, *SE* = 0.91, z = 1.91, p = .06, 95% CI [-0.02, 3.55]). The relationship between question difficulty and recall accuracy does not depend on question content.

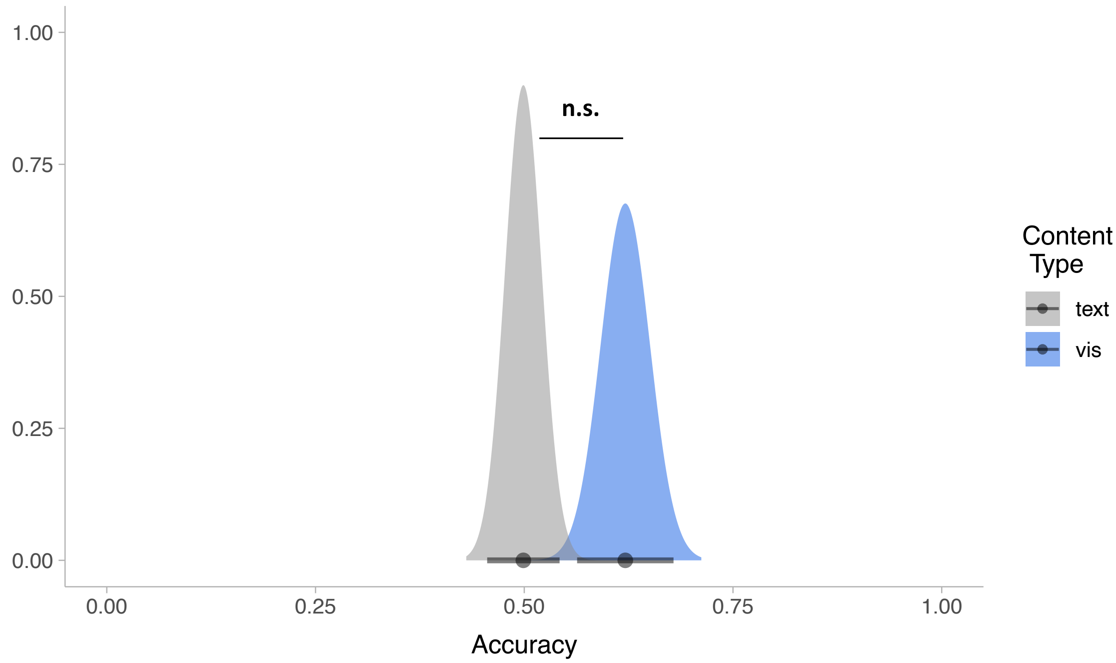

Figure 9. Distribution plot of recall accuracy by question content type. Accuracy is on the x-axis and density is on the y-axis. The plotted distributions are based on standard error. The black lines at the bottom of each curve represent the standard error and each curve represents if the standard error was 95% of the distribution. The black dots represent the means.

Discussion

In Experiment 1, we measured the accuracy, perceived effort and physiological effort via pupil diameter, of 72 participants reading the IPCC report and answering twenty questions. In contrast to our initial pilot, we could not conclude an effect of information format on recall accuracy (H1). This suggests that we could not determine whether there was a difference in recall efficiency between the social media format and the PDF format. For perceived effort, we found that participants reported lower cognitive processing effort for the social media feed format compared to the PDF, which is in line with the results of our pilot work (H2). This means that people perceive themselves as using less cognitive resources when processing information for the social media feed compared to the PDF. Further, we found that participants exhibited greater physiological cognitive processing effort for the social media format compared to the PDF (H3). Therefore, people may prefer social media as an information format even when there is no benefit in terms of memory performance or processing efficiency. This suggests a disconnect between what people perceive in terms of the effort they perceive themselves to be exerting and the resource demand imposed by the two visual formats.

Exploratory Analyses. We were also interested in the content of our test questions. Questions could differ in their presentation, (e.g., the number of words in their chunk or whether they were referring to a visualization) or their difficulty. As expected, when baseline accuracy was higher, recall accuracy also increased. We also found that participants had similar recall accuracy for questions referring to visualizations and questions referring to text. Therefore, there may not be a benefit to presenting information as a visualization or text. This result contrasts with a study by Castro et al. (2022) that found participants reported greater mental demand for textual representations than graphical representations of uncertainty in a decision making task. However, as this is an exploratory analysis, we did not balance the number of questions referring to text and the questions referring to visualizations in our experimental design. As a result, we may see a discrepancy between our findings and previous work.

Limitations. One limitation of this study is the lack of luminance control for the study phase compared to the test phase. We did not control for the varying luminance levels for reading excerpts of the report, which affects pupil size (Peysakhovich et al., 2017). Therefore, some of the variance between the study phase and the test phase can be attributed to differences in the baseline luminance values of the stimuli. However, the test phase is identical across formats, so we are able to compare between them.

Experiment 2

After conducting Experiment 1, we created a version of the study with paper (printed) versions of the displays in order to understand how both format and medium affect the recall and processing of information. As physical PDFs are designed to be read in the physical world, we hypothesized that participants would have higher recall accuracy and lower perceived effort for the PDF compared to the scrolling feed.

Method

Participants. We collected 79 students via the university participant pool and excluded 7 participants due to ignoring task instructions or incomplete responses, leaving 72 participants (17 male, 53 female, 2 non-binary/ third gender). The average age was 21 years (*SD* = 2.08 years).

Stimuli and apparatus. Eye tracking apparatus was identical to Experiment 1. The presented stimuli were identical to Experiment 1 except the reading sections were printed onto paper. The printed readings were presented to the participants using a booklet of bound sheet protectors.

Procedure. The procedure was identical to the Experiment 1 except that the participants only had their eye movements and pupil size recorded as they completed the memory test.

Measures. All measures were identical to the Experiment 1.

Analysis. We first conducted a within-subjects analysis that matched Experiment 1. We utilized a binomial general linear mixed effects model to determine the relationship between information format and recall accuracy. We specified the fixed effect as presentation format and random intercepts were specified for participants (e.g., 1-72) and question number (e.g., questions 1-10). We also implemented a general linear mixed effects model to determine if the information format predicted perceived effort. Our model specified the fixed effect as format and the random intercepts for participants.

We then conducted a between-subjects analysis to investigate the effect of medium on recall accuracy using a binomial general linear mixed effects model. We specified medium and format as fixed effects. We included random intercepts for participants and questions. We also used a general linear mixed effects model to determine the relationship between medium on perceived effort with format, reading section, and medium as fixed effects and random intercepts for participants.

Results

Accuracy. For the physical medium, we could not conclude a difference in accuracy for multiple-choice recall question responses between the scrolling feed format and the PDF format (b = -0.04, SE = 0.11, z = -0.42, p = .67, 95% CI [-0.28, 0.18]).

For the between-subjects analysis, we could not conclude a difference in participant recall accuracy between the two media (b = 0.15, SE = 0.11, z = 1.32, p = .19, 95% CI [-0.08, 0.39]) or the two formats (b = 0.03, SE = 0.08, z = 0.46, p = .65, 95% CI [-0.13, 0.20]).

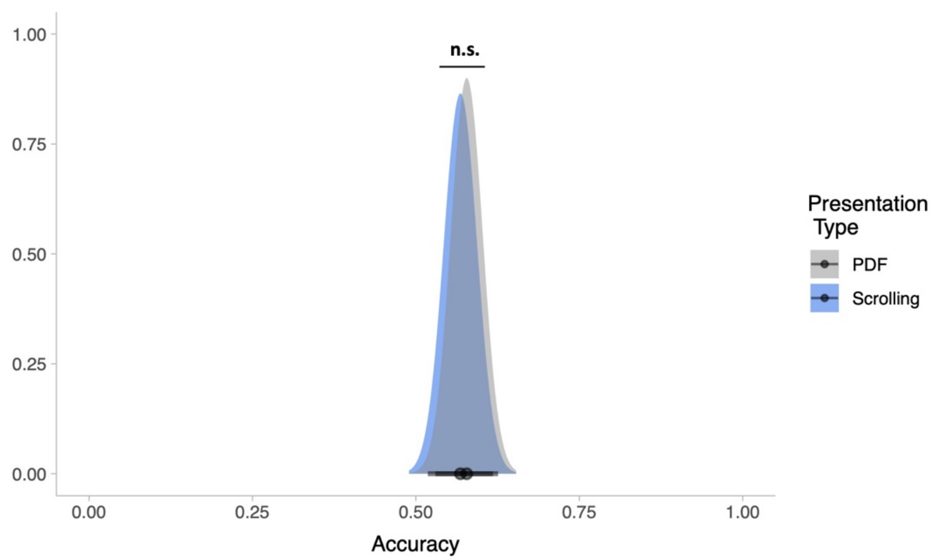

Figure 10. Distribution plot of recall accuracy by format. Accuracy is on the x-axis and density is on the y-axis. The plotted distributions are based on standard error. The black lines at the bottom of each curve represent the standard error and each curve represents if the standard error was 95% of the distribution. The black dots represent the means.

Perceived effort. For the physical medium, participants reported lower perceived effort for the scrolling feed compared to the PDF (b = -1.45, SE = 0.27, t = -5.21, p < .001, 95% CI [-1.99, -0.90]). This effect was significant for three of the NASA Task Load Index Sub-Categories: Mental (How mentally demanding was the task?; b = -2.64, SE = 0.54, t = -4.86, p < .001, 95% CI [-3.71, -1.57]), Frustration (How discouraged or stressed did you feel during the task?; b = -2.35, SE = 0.53, t = -4.43, p < .001, 95% CI [-

3.39, -1.30]), and Effort (How hard did you work to accomplish the task?; *b* = -1.94, *SE* = 0.56, *t* = -3.45, *p* < .001, 95% CI [-3.06, -0.83]).

For the between-subjects analysis, participants reported lower perceived effort for the physical medium compared to the digital medium (*b* = -1.08, *SE* = 0.45, *t* = -2.39, *p* = .02, 95% CI [-1.98, -0.19]) and for the scrolling feed compared to the PDF (*b* = -1.35, *SE* = 0.20, *t* = -6.91, *p* < .001, 95% CI [-1.74, -0.97]). The effect of format was significant for four of the NASA Task Load Index Sub-Categories: Physical (*b* = -0.80, *SE* = 0.24, *t* = -3.30, *p* = .001, 95% CI [-1.28, -0.33]), Mental (*b* = -2.34, *SE* = 0.35, *t* = -6.68, *p* < .001), Frustration (*b* = -2.31, *SE* = 0.39, *t* = -5.99, *p* < .001, 95% CI [-3.03, -1.66]), and Effort (How hard did you work to accomplish the task?; *b* = -1.63, *SE* = 0.40, *t* = -4.09, *p* < .001, 95% CI [-2.41, -0.85]). The effect of medium was significant for two of the NASA Task Load Index Sub-Categories: Performance (How successful were you in accomplishing what you were asked to do?; this scale is reverse coded so that higher values reflect worse performance; *b* = -1.56, *SE* = 0.65, *t* = -2.39, *p* = .02, 95% CI [-2.83, -0.28]) and Mental (*b* = -1.40, *SE* = 0.70, *t* = -1.99, *p* = .048, 95% CI [-2.77, -0.02]).

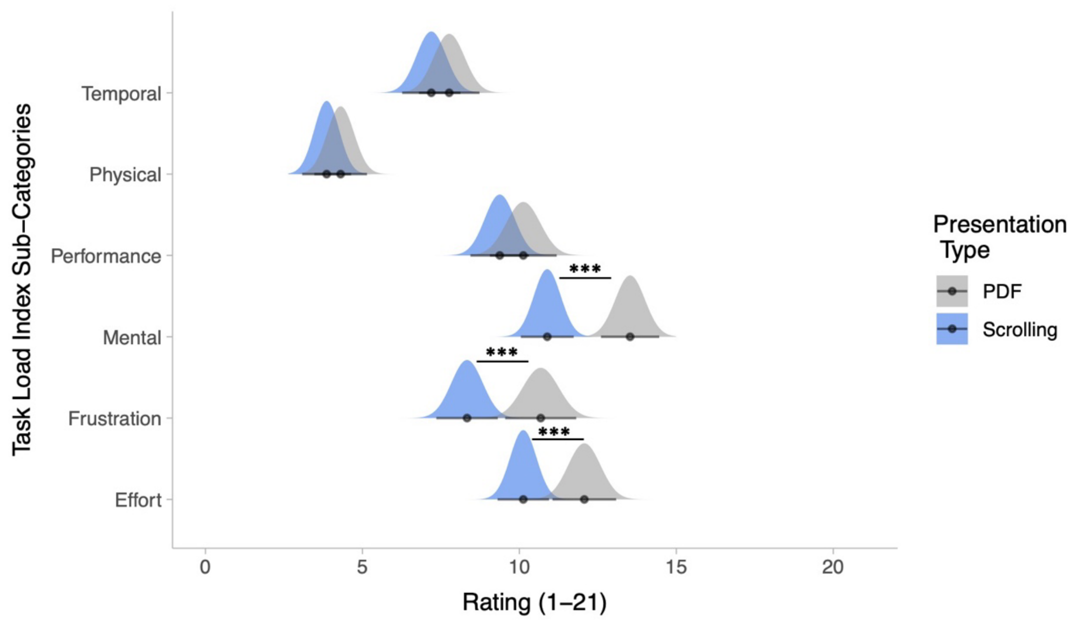

Figure 11. Distribution plot of perceived effort ratings by format, separated by NASA Task Load Sub-Categories. Rating is on the x-axis and sub-category is on the y-axis. The plotted distributions are based on standard error. The black lines at the bottom of each curve represent the standard error and each curve represents if the standard error was 95% of the distribution. The black dots represent the means.

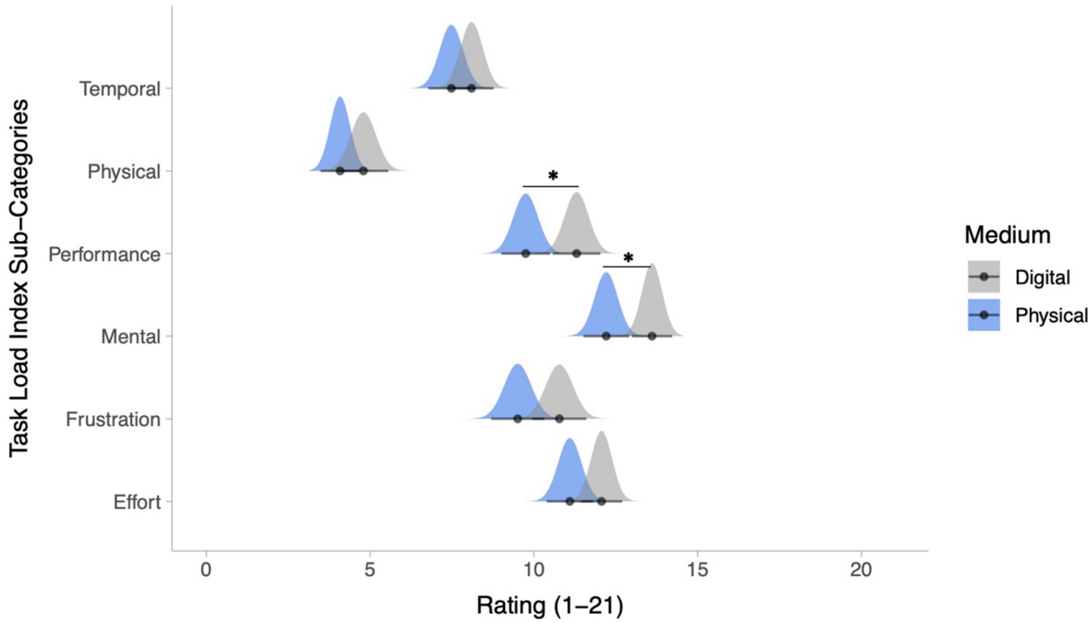

Figure 12. Distribution plot of perceived effort ratings by medium, separated by NASA Task Load Sub-Categories. Rating is on the x-axis and sub-category is on the y-axis. The plotted distributions are based on standard error. The black lines at the bottom of each curve represent the standard error and each curve represents if the standard error was 95% of the distribution. The black dots represent the means.

Physiological effort. For the physical medium, participants had similar mean pupil size (i.e., cognitive effort) during the test phase for the scrolling feed and PDF formats (b = 3.54, SE = 5.00, t = 0.71, p = .48, 95% CI [-6.33, 13.41]).

For the between-subjects analysis, participants also had similar mean pupil size during the test phase for the scrolling feed and PDF formats (b = 6.55, SE = 3.57, t = 1.83, p = .07, 95% CI [-0.47, 13.57]). We also could not conclude a difference in mean pupil size during the test phase for the digital and physical media (b = 44.89, SE = 28.17, t = 1.59, p = .11, 95% CI [-10.32, 100.08]).

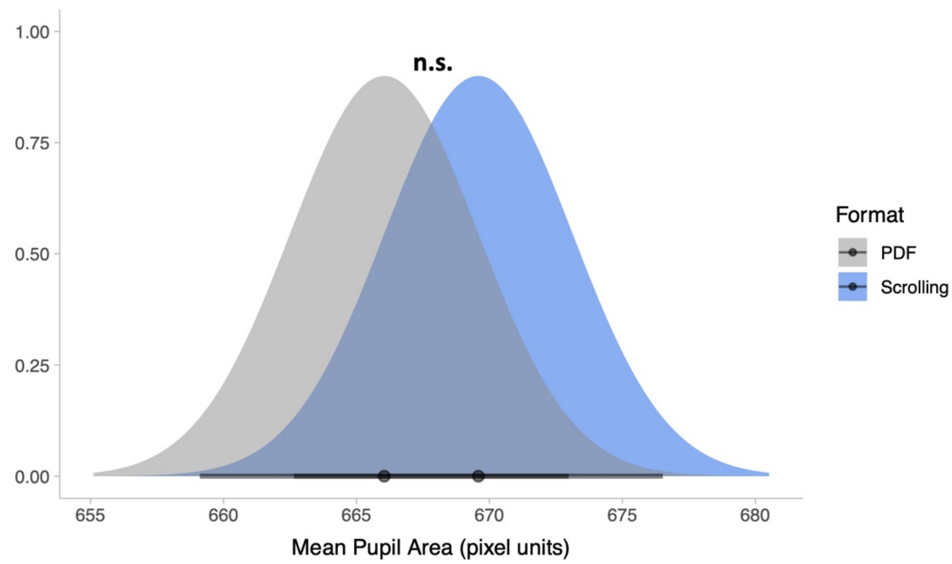

Figure 13. Distribution plot of mean pupil size by format. Mean pupil size is on the x-axis and density is on the y-axis. The plotted distributions are based on standard error. The black lines at the bottom of each curve represent the standard error and each curve represents if the standard error was 95% of the distribution. The black dots represent the means.

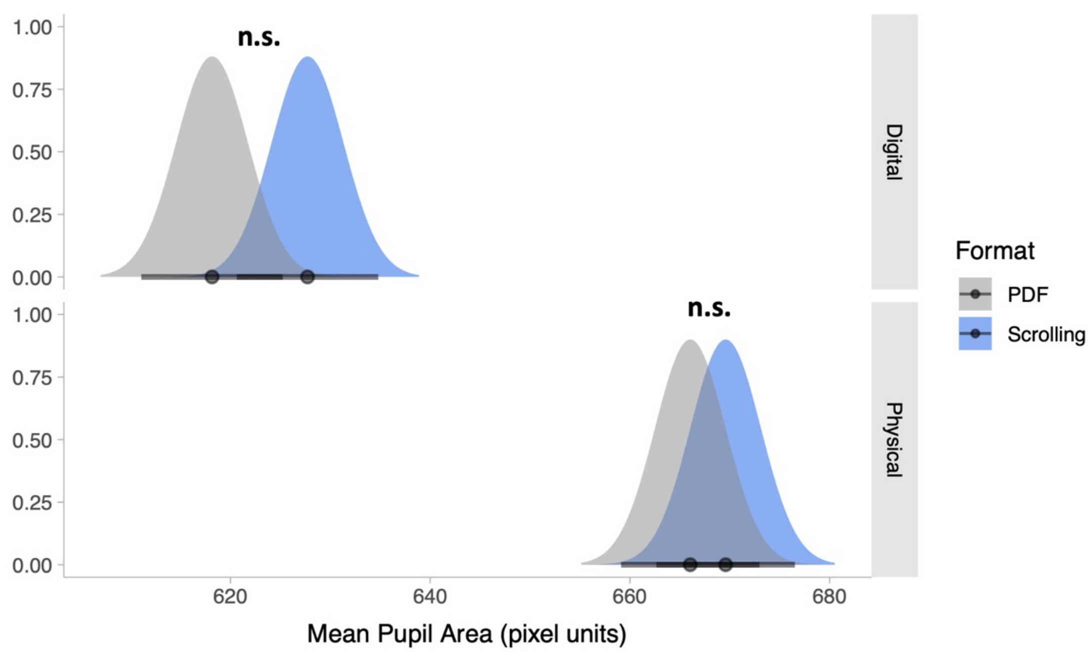

Figure 14. Distribution plot of mean pupil size by format, separated by medium. Mean pupil size is on the x-axis and density is on the y-axis. The plotted distributions are based on standard error. The black lines at the bottom of each curve represent the standard error and each curve represents if the standard error was 95% of the distribution. The black dots represent the means.

Exploratory Analyses. For the physical medium, we found a positive relationship between baseline accuracy and response accuracy ($b = 4.66$, $SE = 0.55$, $z = 8.45$, $p < .001$, 95% CI [3.60, 5.77]). However, after we accounted for question difficulty, we did not find a statistically significant relationship between question content and recall accuracy ($b = -0.75$, $SE = 0.59$, $z = -1.27$, $p = .21$, 95% CI [-1.93, 0.40]). We also did not find an interaction effect between question content and question difficulty ($b = 1.51$, $SE = 0.89$, $z = 1.71$, $p = .09$, 95% CI [-0.20, 3.27]).

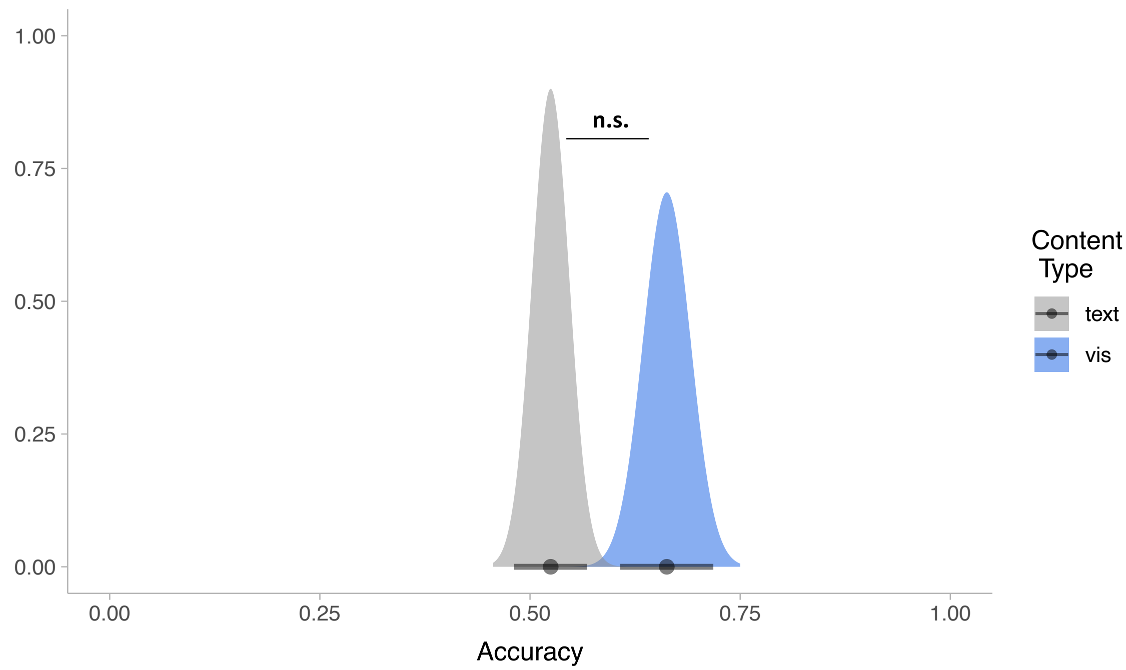

Figure 15. Distribution plot of recall accuracy by question content type. Accuracy is on the x-axis and density is on the y-axis. The plotted distributions are based on standard error. The black lines at the bottom of each curve represent the standard error and each curve represents if the standard error was 95% of the distribution. The black dots represent the means.

Discussion

Similar to Experiment 1, we could not conclude a difference between the two formats for recall accuracy (H1). Despite the change to the physical medium, we do not see an advantage to presenting information in the format of a PDF as we originally hypothesized. Participants had lower perceived processing for the scrolling feed compared to the PDF (H2). In addition, participants reported lower processing effort for the physical medium compared to the digital medium. We could not conclude a difference in physiological effort between the two formats (H3) or the two media.

Exploratory Analyses. For the question characteristics analysis, we see similar results to Experiment 1. We did not find a significant effect of question content on recall accuracy after controlling for question difficulty using baseline average accuracy. In the physical medium, there does not appear to be an advantage in presenting the information as either text or visualization for recall efficiency.

Limitations. One possible explanation for not seeing an advantage in recall accuracy for the PDF in the physical medium is that our participants are mostly young adults in their 20s who are accustomed to seeing PDFs in digital form even when they were intended to be printed out.

Cross-Study Comparisons

Accuracy

Table 1
Cross-Study Comparison Accuracy Score Descriptive Statistics

Study	*M*	*SD*
Online Pilot	0.65	0.16
Experiment 1	0.54	0.14
Experiment 2	0.57	0.15

Participants of the online pilot study had the highest mean recall accuracy (M = 0.65, SD = 016), followed by Experiment 2 (M = 0.57, SD = 0.15) and Experiment 1 (M = 0.54, SD = 0.14). This difference in mean recall accuracy between the online and in-person studies may reflect a discrepancy in the levels of motivation between population groups and therefore differences in the amount of effort dedicated to completing the task.

Perceived Effort

Table 2
Cross-Study Comparison Perceived Effort Descriptive Statistics

Study	*M*	*SD*
Online Pilot	12.08	6.14
Experiment 1	10.11	5.51
Experiment 2	9.02	5.56

Participants of the online pilot had the highest overall mean NASA-TLX score ($M = 12.08$, $SD = 6.14$), followed by Experiment 1 ($M = 10.11$, $SD = 5.51$) and Experiment 2 ($M = 9.02$, $SD = 5.56$). The decrease in NASA-TLX scores between the online version and the in-person versions of the study suggests that the student population used less effort in completing the task compared to the Prolific population.

General Discussion

Across a pilot study and two main experiments, we compared participant's perceived effort, physiological effort, and recall performance between two visual formats and two media. In the pilot study and Experiment 1, participants reported lower perceived effort for the scrolling feed format compared to the PDF format. This result aligns with our expectation of an advantage in the ease of processing arising from the "pre-chunking" of information in social media feeds (Thalmann et al., 2019). Surprisingly, participants in Experiment 2 exhibited the same pattern despite the change from a digital medium to a physical medium. This result contrasts with our initial prediction based on cognitive fit theory (Vessey & Galletta, 1991). In terms of medium, participants reported lower perceived effort for the physical medium compared to the digital medium. In terms of effort, participants reported lower perceived effort for the physical medium compared to the digital medium.

Results from Experiment 1 and 2 reveal that there is a disconnect between participants' perceived effort and physiological effort. For both Experiment 1 and Experiment 2, we found similar physiological effort between the two formats even when participants reported differences in perceived effort. Similarly, there was a difference in perceived effort while physiological effort was similar between the two media. This reflects previous work that found divergence between multiple measures of cognitive effort (Matthews et al., 2015). One possible reason for this disconnect is that unlike other studies measuring cognitive effort (e.g. Gevins & Smith, 2003), we did not deliberately manipulate task demand between conditions. As a result, participants may not be able to differentiate between the Mental (assessing mental demand of the task) and Effort (assessing how much effort was put forth by the participant) subscales of the NASA-

TLX. Therefore, participants may be reporting a difference in the amount of effort they are exerting to complete the task, which is reflected in the NASA-TLX scores while task demands do not necessarily change, which is reflected in pupil size measurements.

We also found mixed results for the effect of format on memory performance. In our online pilot study, participants had higher recall accuracy for the scrolling feed format compared to the PDF format. However, we could not conclude a difference in recall accuracy between the two formats for either Experiment 1 or Experiment 2. We also could not conclude a difference in recall accuracy between the digital and physical media. This result aligns with previous research demonstrating similar reading performance for paging (turning pages) compared to scrolling (Dillon, 1992). We may not have observed an advantage of one medium over the other because the ways participants were interacting with the medium of presentation (turning pages in a booklet or scrolling down a digital screen) tend to result in comparable performance. A possible explanation for the disparity between the online and in-person studies is a difference in study compensation. Online participants were compensated with monetary payment while in-person participants were compensated with research participation credits. As a result, there may have been a shift in motivation between the two study populations, which in turn affects performance (Teigen, 1994). This limitation can be further explored with a future study that manipulates motivation by varying the amount of payment.

Another limitation of the present study is the lack of control of information density for the IPCC report excerpts in order to maintain ecological validity. While the unit of information (i.e., a chunk) is more discrete in the scrolling feed format compared to the PDF format, the level of information density for each chunk varies. However, we plan to address this limitation with future analyses that quantify relative information density with word counts, reading level, and presentation method (i.e., text or visualization) for each chunk.

In this study, we used mean pupil dilation for each reading and test section as a measure of physiological cognitive effort. In a future analysis, we will take a more fine-grained approach and examine how mean pupil dilation changes at the question-level. We will also include other eye tracking measures such as saccades, fixations, and blink rate.

Conclusions

We found that physiological cognitive effort was greater for the social media format, perceived cognitive effort was greater for the PDF format, and we could not determine whether there was a statistically significant difference for behavioral performance between the formats. In terms of effort, people preferred scrolling feeds despite no behavioral advantage for either digital or physical mediums. People also preferred physical paper over a digital screens despite no behavioral advantage.

References

Auxier, B., & Anderson, M. (2021). Social media use in 2021. *Pew Research Center*, *1*, 1–4.

Biondi, F. N., Saberi, B., Graf, F., Cort, J., Pillai, P., & Balasingam, B. (2023). Distracted worker: Using pupil size and blink rate to detect cognitive load during manufacturing tasks. *Applied Ergonomics*, *106*, 103867. https://doi.org/10.1016/j.apergo.2022.103867

Bufano, P., Albertini, N., Chiarelli, S., Barone, V., Valleggi, M., Parrini, G., Tognetti, A., Cesari, V., Gemignani, A., & Menicucci, D. (2022). Weakened Sustained Attention and Increased Cognitive Effort after Total Sleep Deprivation: A Virtual Reality Ecological Study. *International Journal of Human–Computer Interaction*, *0*(0), 1–10. https://doi.org/10.1080/10447318.2022.2099047

Castro, S. C., Quinan, P. S., Hosseinpour, H., & Padilla, L. (2022). Examining Effort in 1D Uncertainty Communication Using Individual Differences in Working Memory and NASA-TLX. *IEEE Transactions on Visualization and Computer Graphics*, *28*(1), 411–421. https://doi.org/10.1109/TVCG.2021.3114803

Castro, S. C., Strayer, D. L., Matzke, D., & Heathcote, A. (2019). Cognitive workload measurement and modeling under divided attention. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(6), 826–839. https://doi.org/10.1037/xhp0000638

Chai, W. J., Abd Hamid, A. I., & Abdullah, J. M. (2018). Working Memory From the Psychological and Neurosciences Perspectives: A Review. *Frontiers in Psychology*, *9*, 401. https://doi.org/10.3389/fpsyg.2018.00401

Chen, S., & Epps, J. (2014). Using Task-Induced Pupil Diameter and Blink Rate to Infer Cognitive Load. *Human–Computer Interaction*, *29*(4), 390–413. https://doi.org/10.1080/07370024.2014.892428

Counting characters. (2023). X Developer Platform. https://developer.twitter.com/en/docs/counting-characters

Cowan, N. (1999). An Embedded-Processes Model of working memory. In *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). Cambridge University Press. https://doi.org/10.1017/CBO9781139174909.006

Dillon, A. (1992). Reading from paper versus screens: A critical review of the empirical literature. *Ergonomics*, *35*(10), 1297–1326. https://doi.org/10.1080/00140139208967394

Gevins, A., & Smith, M. E. (2003). Neurophysiological measures of cognitive workload during human-computer interaction. *Theoretical Issues in Ergonomics Science*, *4*(1–2), 113–131. https://doi.org/10.1080/14639220210159717

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology* (Vol. 52, pp. 139–183). Elsevier. https://doi.org/10.1016/S0166-4115(08)62386-9

Holmes, C. A., Newcombe, N. S., & Shipley, T. F. (2018). Move to learn: Integrating spatial information from multiple viewpoints. *Cognition*, *178*, 7–25. https://doi.org/10.1016/j.cognition.2018.05.003

Just, M. A., Carpenter, P. A., & Miyake, A. (2003). Neuroindices of cognitive workload: Neuroimaging, pupillometric and event-related potential studies of brain work. *Theoretical Issues in Ergonomics Science*, *4*(1–2), 56–88. https://doi.org/10.1080/14639220210159735

Kahneman, D. (1973). *Attention and effort*. Prentice-Hall.

Kramer, A. F. (1986). Interaction between Workload and Training: Converging Evidence from Psychophysiology and Performance Measurement. *Proceedings of the Human Factors Society Annual Meeting*, *30*(11), 1137–1141. https://doi.org/10.1177/154193128603001123

Matthews, G., De Winter, J., & Hancock, P. A. (2020). What do subjective workload scales really measure? Operational and representational solutions to divergence of workload measures. *Theoretical Issues in Ergonomics Science*, *21*(4), 369–396. https://doi.org/10.1080/1463922X.2018.1547459

Matthews, G., Reinerman-Jones, L. E., Barber, D. J., & Abich, J. (2015). The Psychometrics of Mental Workload: Multiple Measures Are Sensitive but Divergent. *Human Factors*, *57*(1), 125–143. https://doi.org/10.1177/0018720814539505

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. https://doi.org/10.1037/h0043158

Nielsen Norman Group. (2020). *Avoid PDF for On-Screen Reading*. Nielsen Norman Group. https://www.nngroup.com/articles/avoid-pdf-for-on-screen-reading/

Peterson, D. J., & Berryhill, M. E. (2013). The Gestalt principle of similarity benefits visual working memory. *Psychonomic Bulletin & Review*, *20*(6), 1282–1289. https://doi.org/10.3758/s13423-013-0460-x

Peysakhovich, V., Vachon, F., & Dehais, F. (2017). The impact of luminance on tonic and phasic pupillary responses to sustained cognitive load. *International Journal of Psychophysiology*, *112*, 40–45. https://doi.org/10.1016/j.ijpsycho.2016.12.003

Recarte, M. Á., Pérez, E., Conchillo, Á., & Nunes, L. M. (2008). Mental workload and visual impairment: Differences between pupil, blink, and subjective rating. *The Spanish Journal of Psychology*, *11*(2), 374–385.

Rodriguez, M. G., Gummadi, K., & Schoelkopf, B. (2014). Quantifying Information Overload in Social Media and Its Impact on Social Contagions. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), Article 1. https://doi.org/10.1609/icwsm.v8i1.14549

Shearer, E., & Mitchell, A. (2021). *News use across social media platforms in 2020*. https://apo.org.au/node/311092

Simon, H. A. (1974). How Big Is a Chunk? *Science*, *183*(4124), 482–488.

SR Research. (2017). *EyeLink Portable Duo User Manual v.1.0.2*. SR Research Ltd., Mississauga, Ontario, Canada.

Teigen, K. H. (1994). Yerkes-Dodson: A law for all seasons. *Theory & Psychology*, *4*(4), 525–547.

Thalmann, M., Souza, A. S., & Oberauer, K. (2019). How does chunking help working memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*, 37–55. https://doi.org/10.1037/xlm0000578

Thomson, K. S., & Oppenheimer, D. M. (2022). The "Effort Elephant" in the Room: What Is Effort, Anyway? *Perspectives on Psychological Science*, *17*(6), 1633–1652. https://doi.org/10.1177/17456916211064896

Vessey, I., & Galletta, D. (1991). Cognitive Fit: An Empirical Study of Information Acquisition. *Information Systems Research*, *2*(1), 63–84.

Voramontri, D., & Klieb, L. (2019). Impact of social media on consumer behaviour. *International Journal of Information and Decision Sciences*, *11*(3), 209. https://doi.org/10.1504/IJIDS.2019.101994

Appendix A

Multiple-Choice Recall Question Set

B1- (B): Why are five emission scenarios being considered in the report?

A. To consider the climate response to the broader range of greenhouse gas
B. To consider land use
C. To consider pollutant future
D. **All the above**

B2- (SPM.4a): What are the three key non-carbon dioxide (CO2) drivers of additional warming?

a. **Methane, nitrous oxide, sulfur dioxide**
b. Methane, carbon monoxide, sulfur dioxide
c. Propane, nitrous dioxide, sulfur dioxide
d. Propane, nitrous oxide, copper oxide

B3- (SPM.4a): For all scenarios, future annual emissions of carbon dioxide (CO_2) decrease between 2015-2100.

A. True
B. **False**

B4- (SPM.4b): Which of the following is projected to result in net cooling?

A. Carbon dioxide (CO2)
B. Non-carbon dioxide (CO_2) greenhouse gases
C. **Aerosols and land use**
D. All the above

B5- (SPM.4b): Which of the following emissions contributes the most to the global surface temperature increasing?

a. Non- carbon dioxide (CO_2)greenhouse gases
b. **Carbon dioxide (CO_2)**
c. Aerosols and land use
d. None of the above

B6- (B.1): Global surface temperature will increase until at least what century unless reductions in greenhouse gas emissions occur?

a. Early-century
b. **Mid-century**
c. Late-century
d. Next century

B7- (B.1.1): When was the last time the global surface temperature was over 2.5°C?

a. **Over 3 million years ago**

b. 1 million years ago
c. 300 years ago
d. 30 years ago

B8- (B1.2): Under the five illustrative scenarios, global warming of 5°C is exceeded during the 21st century under the intermediate greenhouse gas emission scenario.

A. True
B. **False**

B9- (B1.3): Under the five illustrative scenarios, in the near term (2021-2040) the 1.5°C global warming level is very likely to be exceeded under the very high greenhouse gas emission scenario.

a. **True**
b. False

B10- (B1.4): How can global surface temperature vary due to natural variability in relation to long-term human-induced trends?

A. Above
B. Below
C. **Both**
D. Neither

D1- (D.1)

Reducing methane (CH_4) emissions would limit the warming effect by declining aerosol pollution and which of the following?

a. Declining CO_2 emissions
b. **Improving air quality**
c. Increasing CO_2 emissions
d. Decreasing air quality

D2- (D.1.1)

This report reaffirms these emissions cause an average increase of which of the following in global surface temperature?

a. 0.35°C
b. **0.45°C**
c. 0.55°C
d. 0.65°C

D3- (D.1.1)

To stabilize human-induced global warming, anthropogenic carbon dioxide (CO_2) emissions have to reach which of the following?

a. **Net zero**
b. Below zero
c. Above zero
d. Near zero

D4- (SPM.10)

Cumulative carbon dioxide (CO_2) emissions are related to ____ in global surface temperature.

A. Decrease
B. **Increase**
C. No change
D. Minor change

D5- (SPM.10)

The relationship between cumulative carbon dioxide (CO_2) emissions and projected global warming is:

A. Near exponential
B. Near logarithmic
C. **Near linear**
D. Near constant

D6- (SPM.10)

Future cumulative carbon dioxide (CO_2) emissions are the same across scenarios.

A. True
B. **False**

D7- (D.1.2)

Over the period of 1850-2019, the remaining carbon dioxide (CO_2) budgets were estimated from which of the following factors?

A. CO_2 emissions only
B. Methane (CH_4) emissions only
C. Regional temperature limits
D. **Global temperature limits**

D8- (D.1.2)

Which of the following was also included to estimate carbon dioxide (CO_2) budgets?

a. Feedback from methane (CH_4) emissions
b. CH_4 emissions reaching net zero
c. **Projection of warming from non-CO_2 emission**
d. Rising sea levels

D9- (D.1.3)

Factors determining the estimates of remaining carbon budget have been reassessed and are similar to the magnitude of which of the following?

A. **SR 1.5**
B. SR 2.5
C. SR 1.0
D. SR 1.3

D10- (D.1.4)

Anthropogenic carbon dioxide removal (CDR) aims to reach the goal of what?

A. Net zero carbon dioxide (CO_2) emissions
B. Net zero greenhouse gas emissions
C. Lower surface temperature
D. **All the above**