Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2025) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

29th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2025)

The Disfluency Effect in Reading Comprehension: Findings from Paper and Screen Experiments

Ai Hagihara[a], Yuhi Sezaki[a], Sayaka Takano[a], Satoshi Nakamura[a], Akiyuki Kake[b]

[a]*Meiji University, 4-21-1 Nakano, Nakano-ku, Tokyo 164-8525, Japan*
[b]*Wacom Co., Ltd., 2-510-1 Toyonodai, Kazo-shi, Saitama 349-1148, Japan*

Abstract

Numerous studies have investigated the relationship between text format and memory performance, revealing that less readable text formats can sometimes enhance memory retention. This counterintuitive phenomenon is often referred to as the disfluency effect, which suggests that perceptual difficulty may lead to deeper cognitive processing. The objective of this study was to explore whether a similar disfluency effect would occur in reading comprehension. We therefore conducted two experiments: a paper-based experiment and a screen-based experiment. In both experiments, participants read passages presented in two font styles and two types of handwritten text, followed by reading comprehension questions. In the paper-based experiment, participants tended to perform slightly better when the text was more readable, although the difference was not statistically significant. In the screen-based experiment, where response times were recorded, participants spent less time answering when the text was less readable, and their accuracy was also slightly lower. These findings suggest that, in the case of long passages such as those used in this study, the cognitive load imposed by less readable text may interfere with comprehension, potentially offsetting the benefits predicted by the disfluency effect.

Keywords: handwritten text; font; text format; reading comprehension questions; comprehension accuracy;

1. Introduction

The proliferation of electronic devices such as PCs, smartphones, and tablets has transformed learning methods, increasing opportunities for digital materials and typed note-taking. Nevertheless, handwritten note-taking remains prevalent, particularly in primary education. Consequently, numerous studies have investigated the effects of handwritten note-taking compared to typing. Mueller et al. [7] found that students who took notes using a laptop performed worse on conceptual questions than those who took notes by hand. Similarly, Mendizábal et al. [2] reported that students who took notes by hand outperformed those who used a computer on memory tests.

In addition, studies comparing the characteristics of paper, iPads, and PCs as learning materials have revealed that paper tends to excel in fundamental and knowledge-based tasks, while iPads often demonstrate superior performance in applied and integrative tasks. As a result, combining paper and iPads is considered to yield the most effective

learning outcomes [1]. Given that modern learning environments integrate both digital and analogue elements, understanding which visual textual features such as font style and character shape are most effective for learning is crucial. Moreover, according to Bloom [3], learning is known to progress through a hierarchical sequence of stages: remembering, understanding, applying, analyzing, evaluating, and creating. Each stage requires the achievement of the objectives of the preceding stage. Consequently, the stages of remembering and understanding are particularly important as they form the foundation for subsequent stages of learning.

We have previously conducted research on the effects of font styles and handwritten characters on memory when using Japanese text. In our memory experiment, we utilized two types of fonts and two types of handwritten characters with distinct features. The results showed that memory scores decreased when text was presented in a Gothic-style font (a type of font similar to sans-serif font commonly used in Japanese) [6]. In contrast, for handwritten characters, scores were higher when the handwriting was less legible [11]. This result aligns with findings by Oppenheimer et al. [4] and Sungkhasettee et al. [10], who demonstrated that presenting text in a difficult-to-read format can enhance memory. This phenomenon, where imposing a moderate cognitive load during information processing, such as presenting text in a less legible font, can improve memory, is known as the disfluency effect.

While many studies have examined how text readability influences memory, fewer have focused on whether similar effects occur in reading comprehension. In order to explore this question, several studies have investigated the relationship between text format and comprehension. Dressler [5] found that comprehension improved when texts were presented in an easy-to-read font or without time constraints. Zhou et al. [13] showed that font style and color affected users' cognitive load in reading news articles, suggesting that visual presentation influences information processing. In Japanese, Negishi et al. [8] examined the effect of font style on comprehension, but did not compare fonts with handwritten text. Therefore, it remains unclear whether fonts or handwritten characters are more effective for reading comprehension, underscoring the need for further investigation.

In this study, we aim to investigate whether the disfluency effect observed in memory performance also applies to reading comprehension. Specifically, we focus on how the legibility of text, which is manipulated through font styles and handwritten character types, affects comprehension accuracy. To this end, we conducted two experiments. One was in a paper-based (analogue) environment, and the other was in a screen-based (digital) environment. In both experiments, participants answered reading comprehension questions based on passages presented in four text conditions. These conditions consisted of two fonts and two types of handwritten text. The texts used in the experiments were written in Japanese, the native language of all participants. This was done to control for language familiarity and isolate the effect of visual readability.

This two-experiment design allowed us to examine the effects of text readability across different media. We hypothesized that, as in memory tasks, lower legibility might improve performance by encouraging deeper processing. However, the increased cognitive load associated with less readable text might instead hinder comprehension, especially when reading longer passages.

The main contributions of this study are as follows:

- We extend the investigation of the disfluency effect from memory performance to reading comprehension, which involves more complex cognitive processing.
- We compare the impact of both font styles and handwritten characters on comprehension, a combination rarely explored in previous research.
- We observed a tendency for less readable character formats to result in shorter response times in comprehension tasks. This trend suggests that character readability may influence the depth of text processing, indicating that character shape could be a relevant factor in future research and educational practice.

2. Related Work

Numerous studies have investigated the disfluency effect in learning and memory. Diemand-Yauman et al. [4] examined the influence of font readability on academic performance by distributing learning materials printed in either easy-to-read or less readable text. Their study found that students who studied with less readable fonts scored significantly higher on subsequent tests. Seufert et al. [9] manipulated font type, size, and color to create four levels of readability (Fluent, Disfluency 1–3), and conducted content comprehension tests. The results showed that the

Disfluency 1 condition yielded the highest scores, whereas Disfluency 3 resulted in the lowest scores. These findings suggest that a moderate level of disfluency enhances comprehension, whereas excessive disfluency may hinder it. Additionally, Sungkhasettee et al. [10] investigated the learning effects of presenting words upside down. They found that participants had a higher recall rate for inverted words, despite being unaware of the memory advantage this format provided.

While these studies focus on disfluency effects in English, similar research has been conducted in Japanese. Tanigami [12] examined the impact of text style on readability and memory using a sample of ten university students. Passages were presented in three different styles (textbook font, Gyoseki Gyosho font, and handwritten text), and word recall tests were administered. The results showed no significant differences in memory performance across the styles. In a five-point questionnaire evaluating comprehensibility, no effects of text style on perceived understanding were observed. In contrast, Negishi et al. [8] demonstrated the effects of font type on reading comprehension and memory. Participants read passages in easy-to-read fonts (MS Gothic, MS Mincho) and hard-to-read fonts (HGS textbook font, HGS Gyosho font), followed by an interval and a series of comprehension and fill-in-the-blank tasks. The results indicated that participants tended to perform better in both comprehension and recall when the text was presented in a less readable font.

As illustrated above, numerous studies have examined the relationship between text format and learning outcomes. However, the findings remain inconsistent, highlighting the need for further research. In particular, while many studies have focused on memory, fewer have examined comprehension, especially with regard to differences between handwritten and computer-rendered text. Therefore, this study aims to investigate the impact of text format differences, specifically between handwritten and printed texts, on reading comprehension.

3. Experimental Design: Reading Comprehension Tasks

3.1. Experimental Overview

This experiment investigates how differences in text format, including fonts and handwritten styles, affect reading comprehension. It tests the hypothesis that presenting reading passages in less legible text may improve comprehension accuracy. The text formats used in the experiment included two types of fonts and two types of handwritten text with distinct characteristics. Participants were presented with texts written in each of these formats and asked to answer comprehension questions in a true-or-false format. To examine the influence of different presentation media, we conducted the experiment in both paper-based and screen-based conditions.

3.2. Text Format Selection: Fonts and Handwriting Styles

This section describes the selection of two fonts and two types of handwritten text, along with the rationale for their choice. For the font formats, MS Mincho and MS Gothic were selected based on their use in previous memory experiments [6]. As for the handwritten text, two distinct types of handwriting were chosen. These samples were selected from a dataset collected in previous research [11], which aimed to assess handwriting similarity.

Among the individuals who agreed to handwrite the experimental text (a total of 9,530 characters), we selected two whose handwriting had the lowest cosine similarity score, calculated based on a handwriting impression evaluation. This selection ensured that the visual styles of their handwriting were clearly distinct. The selected handwriting samples and their characteristics are shown in Fig. 1. As shown in the figure, Handwriting 1 features a cute and rounded style, while Handwriting 2 exhibits a mature and bold cursive style.

3.3. Task Design

In designing the task for this experiment on reading comprehension, we considered various types of questions, including trick questions, math problems with misleading elements, self-created questions based on newspaper articles, and SPI (Synthetic Personality Inventory) reading comprehension questions. However, we identified several limitations in using these alternatives. Trick questions and math problems tend to involve relatively short passages, making them inadequate for assessing participants' understanding of longer texts. Newspaper articles, on the other

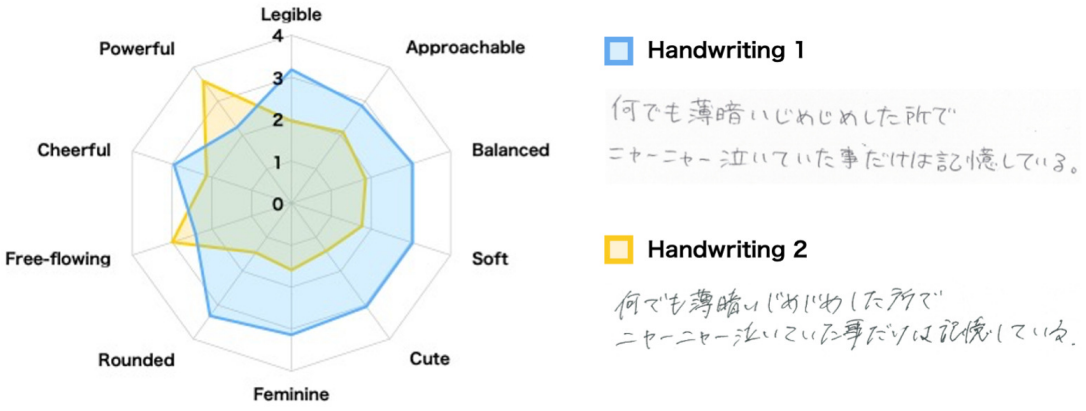

Fig. 1. Characteristics and examples of handwritten text used in the experiment

hand, often cover themes unfamiliar to university students, the participants in this study, potentially hindering comprehension due to a lack of background knowledge. Additionally, self-created questions may contain flaws in their structure or fail to appropriately measure comprehension due to insufficiently calibrated difficulty levels. Based on these considerations, we decided to use SPI reading comprehension practice questions, which are specifically designed to assess the ability to logically interpret written content. We deemed these questions suitable for effectively evaluating participants' comprehension in this experiment.

Furthermore, if the correct answer is explicitly stated in the passage, participants may succeed by simply locating the matching text without fully understanding the content. To address this concern, we selected questions in which the correct answer is not directly stated in the text, ensuring that participants must rely on comprehension rather than simple text matching.

To control for response strategies and accurately measure comprehension, we established the following guidelines:

- Participants were instructed not to read the questions before reading the passage.
- They were required to read the entire passage from start to finish before proceeding to the questions.
- After reading the passage once, participants were allowed to refer back to the text while answering the questions.

These measures were implemented to minimize the influence of various answering techniques and ensure participants engaged in comprehensive reading before responding to the questions.

Additionally, we imposed a time limit on each set of questions, adjusting the limit according to the total word count of the passage and accompanying questions. Specifically, the time limit was set between 3 to 5 minutes per set of questions.

3.4. Question Presentation Method

First, we created questions using the four types of text formats selected in Section 3.2. An example of these questions is shown in Fig. 2. In the paper-based environment, the questions were printed on A4 paper, while in the digital environment, we used a custom experimental system.

The experimental system was designed to control the answer time and collect data on both the time spent reading the passage and the time taken to answer the questions. Participants could freely navigate between the page displaying the reading passage and the combined question-and-answer screen for the true/false task, as shown in Fig. 3. To answer the questions, participants could click within the parentheses provided in the answer field. Initially, the parentheses were left blank; however, a preliminary experiment revealed that many participants were unsure how to interact with

Fig. 2. An example of questions presented in four text formats used in the experiment
(From left to right: MS Gothic, MS Mincho, Handwriting 1, Handwriting 2)

the answer interface. Therefore, we set the default value within the parentheses to "◯" to improve the clarity of the input method and reduce participant confusion.

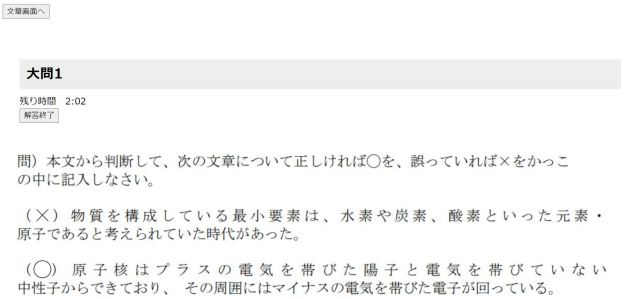

Fig. 3. Answer screen displayed after reading the passage

The system was implemented using PHP, JavaScript, and MySQL. This implementation ensured consistency in question presentation and enabled us to collect detailed behavioral data, such as reading and response times, in the screen-based environment.

4. Experiment 1: Paper-Based Reading Comprehension Task

4.1. Experimental Procedure

36 university students participated in this experiment. Each text format was presented twice per participant across eight question sets to ensure balanced exposure.

The experimental procedure is shown in Fig. 4. Participants were first informed about the answer instructions established in section 3.3. They then completed a total of nine sets of questions: one practice set and eight main sets. The practice set was presented in a different font (Yu Kyokashotai) from those used in the main sets. A time limit of 3

to 5 minutes was set for each question set, depending on the length of the passage. After completing all the questions within a set, participants were asked to respond to a questionnaire regarding the most recent main question. In this questionnaire, participants were asked about the readability of the presented text format and their confidence in their answers.

After completing all sets of questions and their corresponding questionnaires, participants filled out a Google Form survey. This final survey gathered demographic information such as gender, faculty, and the frequency of reading text on electronic devices, printed text, and handwritten text. It also included questions about participants' perceived difficulty with the Japanese subject.

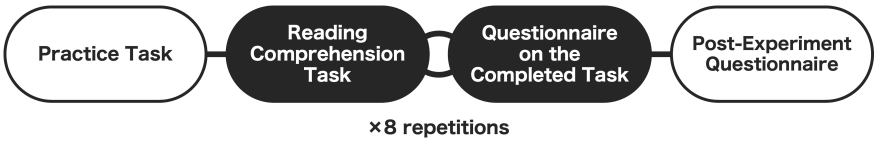

Fig. 4. Experimental procedure in the paper-based environment

4.2. Results

This section describes the analysis results based on the scores of all 36 participants, as no outliers were detected. Table 1 shows the accuracy rates for each main question. Although the initial plan was to conduct a within-subjects comparison, the table reveals considerable variation in scores due to differences in the difficulty levels of each main question. Consequently, we conduct a between-text-format comparison based on overall average scores.

Table 1. Accuracy rate for each main question

Question Number	1	2	3	4	5	6	7	8
Accuracy Rate (%)	80.6	88.9	95.4	93.5	77.8	99.1	100	65.7

Fig. 5 shows the average accuracy rate for each text format, based on the mean score across all question sets. The horizontal axis represents the average accuracy rate. Among the text formats, MS Gothic had the highest average accuracy rate, while Handwriting 2 had the lowest. However, multiple comparisons of the average scores among text formats revealed no significant differences.

Subsequently, Fig. 6 presents the average readability scores for each text format. The horizontal axis represents the average readability rating, measured on a four-point Likert scale, where 4 indicates "very easy to read" and 1 indicates "very difficult to read." The results indicate that MS Mincho had the highest readability score, followed by MS Gothic, Handwriting 1, and Handwriting 2, which had the lowest readability score. Multiple comparisons revealed no significant difference between the two font types; however, significant differences were observed among all other font pairs except between MS Mincho and MS Gothic ($p < 0.01$).

Based on the readability questionnaire, participants who rated each text format as easy to read were classified as the "high readability group," while those who rated it as difficult to read were classified as the "low readability group."

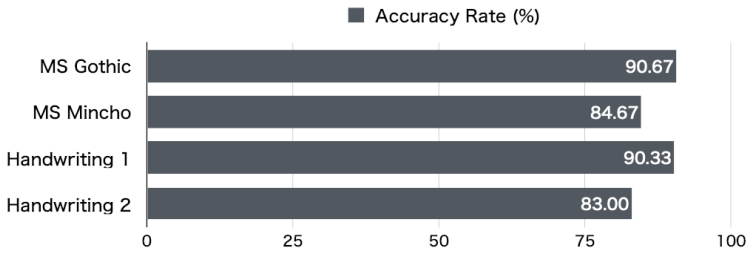

Fig. 5. Average accuracy rate by text format

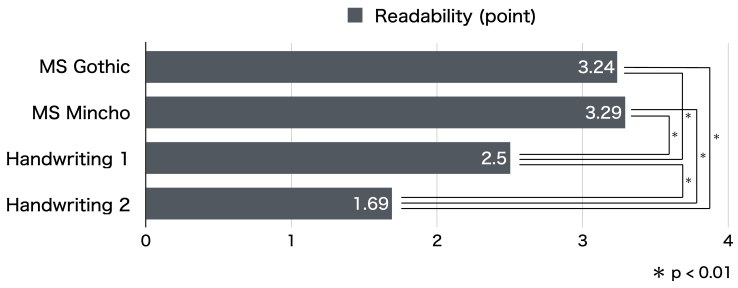

Fig. 6. Average readability rating for each text format

As a result, the average accuracy rate for the high readability group was 89.00%, while that for the low readability group was 84.33%, showing no statistically significant difference between the two groups.

Table 2 presents the distribution of readability ratings for each text format. The results show that the majority of participants rated the fonts as easy to read, while most participants rated Handwriting 1 as difficult to read. Consequently, focusing on the average accuracy rates for Handwriting 1 based on readability ratings, the high readability group achieved an average score of 97.33%, while the low readability group scored 85.67%. The results indicate that participants who rated Handwriting 1 as easy to read achieved higher scores, whereas those who rated it as difficult to read had lower scores.

Table 2. Distribution of ratings for each text format

	MS Gothic	MS Mincho	Handwriting 1	Handwriting 2
Very Easy to Read	22	30	6	1
Easy to Read	48	35	31	6
Difficult to Read	1	5	28	35
Very Difficult to Read	1	2	7	30

4.3. Limitations of the Paper-Based Experiment

When focusing on Handwriting 1, which showed divided readability ratings, we observed that participants who perceived the text as easy to read tended to achieve higher reading comprehension scores. However, in the paper-based experiment, participants were asked to rate the readability of the text immediately after answering each set of questions. As a result, their performance or confidence in their answers may have influenced their readability ratings. To address this potential confound, we plan to revise the questionnaire format in the upcoming screen-based experiment.

Furthermore, the results of the paper-based experiment revealed a bias in readability ratings depending on the text format. Although it is important to investigate how readability affects reading time and answer time, such measurements were difficult to control and record accurately in the paper-based environment. Therefore, in the screen-based experiment to follow, we will also examine reading time and answer time in relation to text readability.

5. Experiment 2: Screen-Based Reading Comprehension Task

5.1. Experimental Procedure

46 university students participated in this experiment. Each participant was assigned to a single text format, and all main question sets were presented in that format. The content of the questions was the same as in the paper-based experiment, except for Question Set 7, which was excluded due to a 100% accuracy rate.

The experimental procedure is illustrated in Fig. 7. Participants first completed one practice question, followed by seven main question sets, for a total of eight sets. The practice question was presented in a different font (Yu Kyokasho) from those used in the main trials. A time limit of 3 to 5 minutes was imposed for each question set, depending on the length of the passage. After completing all the questions, participants answered a post-experiment questionnaire regarding their overall experience, followed by a comparative survey on the readability of each text format.

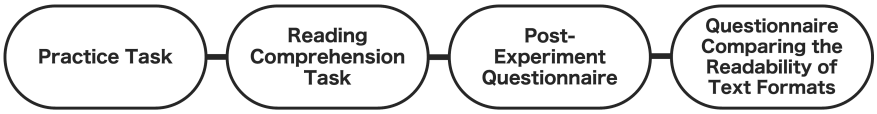

Fig. 7. Experimental procedure in the screen-based environment

5.2. Results

Outliers were identified in both answer times and accuracy scores based on the experimental data. Therefore, data from four participants were excluded from the analysis, and the results of the remaining 42 participants were used. Table 3 presents the accuracy rates for each question set. The table indicates that variations in difficulty levels across question sets led to fluctuations in participants' scores. Accordingly, we conducted comparisons based on overall average scores in this experiment as well.

Table 3. Accuracy rate for each main question

Question Number	1	2	3	4	5	6	8
Accuracy Rate (%)	57.04	74.31	89.21	89.03	66.53	97.22	40.05

Fig. 8 displays the average accuracy rates for each text format, calculated based on the mean score across all question sets. The horizontal axis represents the average accuracy rate. According to the figure, MS Mincho yielded the highest average score, while Handwriting 2 showed the lowest. However, multiple comparisons of the average scores across text formats revealed no statistically significant differences.

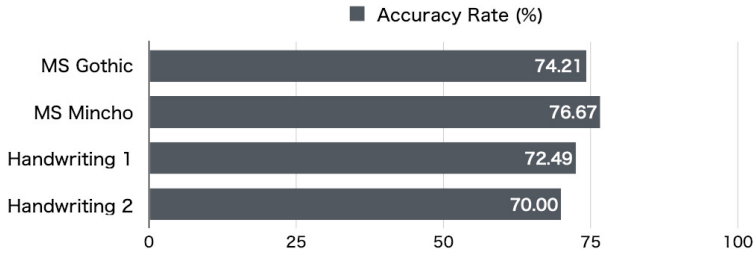

Fig. 8. Average accuracy rate by text format

Fig. 9 shows the average readability ratings for each text format. The vertical axis represents the average values obtained from a four-point Likert scale, where 4 indicates "very easy to read" and 1 indicates "very difficult to read." According to the figure, MS Gothic was rated the most readable, followed by MS Mincho, Handwriting 1, and Handwriting 2, which was rated the least readable. Multiple comparisons revealed significant differences in readability across all text formats.

Based on the readability questionnaire, participants who rated each text format as easy to read were classified into the high readability group, while those who rated it as difficult to read were grouped into the low readability group. A comparison of their performance showed that the high readability group had an average accuracy rate of 76.37%, whereas the low readability group scored 68.57%, indicating a tendency for higher accuracy when the text is perceived as more readable.

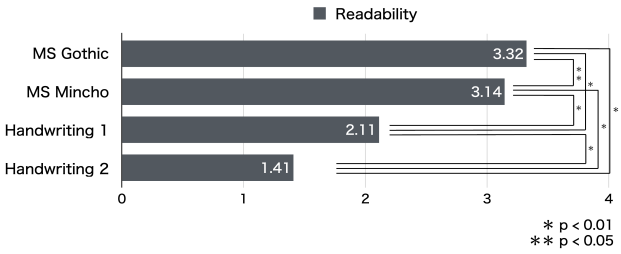

Fig. 9. Average readability rating for each text format

Fig. 10 shows the relationship between the readability of each text format and the corresponding average answer time when that format was presented. The figure indicates a trend in which the more readable the text was perceived to be, the longer participants took to answer. This suggests that when questions are presented in less readable text, participants may be less inclined to read carefully, which may, in turn, contribute to reduced accuracy.

Fig. 11 shows the relationship between the number of correct answers and the average answer time for each text format. The correlation coefficient between the number of correct answers and answer time was 0.31, indicating a weak positive correlation. This correlation suggests that, within the time limit, spending more time on each question was associated with a higher number of correct answers.

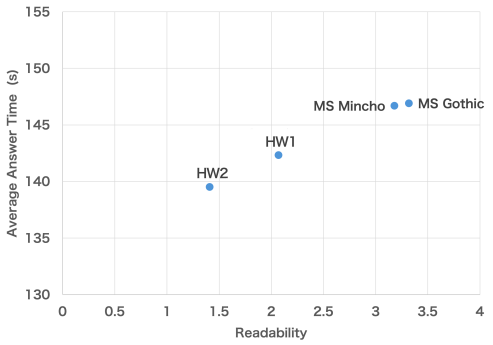

Fig. 10. Relationship between text readability and answer time

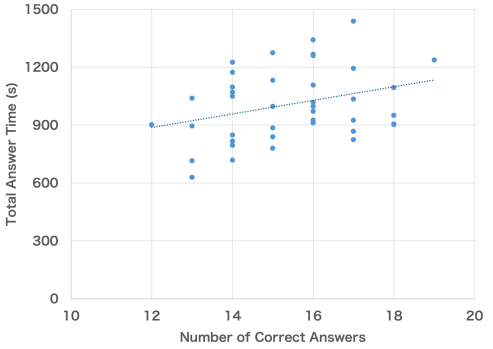

Fig. 11. Relationship between answer time and number of correct answers

6. Discussion

Although the timing of readability evaluations differed between the paper-based and screen-based experiments, both showed a non-significant trend suggesting that participants tended to achieve slightly higher scores when reading more readable text formats. While these differences were not statistically significant, the consistency of the trend across both environments suggests that the benefits of disfluency observed in memory tasks may not directly apply to reading comprehension. In memory studies, less readable text has been shown to enhance retention, whereas in this context, it may hinder comprehension.

In the case of long passages, such as those approximately 1,000 characters in length used in the present study, reduced readability may impose excessive cognitive load, thereby interfering with comprehension processes rather than promoting them. Moreover, the texts used in the experiments were drawn from subjects such as science, social studies, and philosophy. These texts required readers to understand logical relationships and sentence structures in order to answer the questions. These tasks may differ from those involving narrative texts, where understanding emotions or intentions plays a more central role. This remains an area for further investigation.

Additionally, while this study focused on reading comprehension as the primary indicator of understanding, alternative evaluation methods could yield different results. For instance, in tasks involving misleading mathematical

problems with a much smaller volume of text, the disfluency effect might still emerge and support deeper engagement with the content.

Overall, the present findings suggest that the effectiveness of disfluency may depend on the nature of the task as well as the length and complexity of the reading material. Understanding is a multifaceted cognitive process, and different types of comprehension such as literal, inferential, or emotional understanding may be influenced in different ways by variations in text readability. Future research should continue to investigate how these aspects interact. Such insights will be important for improving the design and evaluation of educational materials, especially in digital environments where the visual presentation of text can be adjusted according to the purpose of the learning task.

7. Conclusion

In this study, we tested the hypothesis that presenting reading passages in less readable text formats would improve comprehension accuracy. To examine this hypothesis, we conducted an experiment using two types of fonts and two types of handwritten text. The tasks were administered in two conditions: paper-based and screen-based. The results showed that, in both environments, participants tended to perform slightly better when the text was more readable, although the differences were not statistically significant. However, in the screen-based condition, where answer times were recorded, less readable text tended to result in faster responses and slightly lower comprehension accuracy. These findings suggest that, for long passages such as those used in this experiment, the cognitive load imposed by less readable text may have outweighed the potential benefits of disfluency.

References

[1] Akahori, K., Wada, Y., 2012. Characteristics comparison of ipads, paper and pcs as learning devices. Hakuo Daigaku Kyoiku Gakubu Ronshu (Bulletin of the Faculty of Education, Hakuoh University) 6, 15–34. (in Japanese).

[2] Aragón-Mendizábal, E., Delgado-Casas, C., Navarro-Guzmán, J., Menacho-Jiménez, I., Romero-Oliva, M.F., 2016. A comparative study of handwriting and computer typing in note-taking by university students. Comunicar 24, 101–107.

[3] Bloom, B.S., Engelhart, M.D., Furst, E.J., Hill, W.H., Krathwohl, D.R., et al., 1956. Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain. Longman New York.

[4] Diemand-Yauman, C., Oppenheimer, D., Vaughan, E., 2010. Fortune favors the: Effects of disfluency on educational outcomes. Cognition 118, 111–5. doi:10.1016/j.cognition.2010.09.012.

[5] Dressler, E., 2019. Understanding the effect of font type on reading comprehension/memory under time-constraints. Theses/Capstones/Creative Projects 70 .

[6] Ito, R., Hamano, K., Nonaka, K., Sugano, I., Nakamura, S., Kake, A., Ishimaru, K., 2020. Comparison of the remembering ability by the difference between handwriting and typeface. Technical Report. HCI International 2020-Posters: 22nd International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part I 22.

[7] Mueller, P.A., Oppenheimer, D.M., 2014. The pen is mightier than the keyboard: Advantages of longhand over laptop note taking. Psychological science 25, 1159–1168.

[8] Negishi, I., Kodama, M., 2018. Influence of japanese font type on learning effectiveness. Kogaku Kyoiku (Japanese Journal of Engineering Education) 66, 48–412. (in Japanese).

[9] Seufert, T., Wagner, F., Westphal, J., 2017. The effects of different levels of disfluency on learning outcomes and cognitive load. Instructional Science 45, 221–238.

[10] Sungkhasettee, V.W., Friedman, M.C., Castel, A.D., 2011. Memory and metamemory for inverted words: Illusions of competency and desirable difficulties. Psychonomic bulletin & review 18, 973–978.

[11] Takano, S., Yamazaki, I., Ito, R., Hamano, K., Sugano, I., Nakamura, S., Kake, A., Ishimaru, K., 2022. Examining the effect of similarity to one's own handwriting on ease of memory. Technical Report 2. Kenkyu Hokoku Human Computer Interaction (HCI). (in Japanese).

[12] Tanigami, A., 2019. The effects of font styles or handwriting on the fluency of reading texts and memory. Hikone Ronso , 18–29(in Japanese).

[13] Zhou, J., Miao, X., He, F., Miao, Y., 2022. Effects of font style and font color in news text on user cognitive load in intelligent user interfaces. IEEE Access 10, 10719–10730. doi:10.1109/ACCESS.2022.3151915.