

Beyond Words

Exploring the Effects of Font Size & Pacing on Reading Enjoyment and Fact Recall

John Gibbons, Andre Gigena, Maria Manna, and Ethan Moody

December 2023

Abstract

Existing research suggests that font and text characteristics can impact the way an individual engages with a reading experience. This study investigates the effect of font size and forced-pacing on reading enjoyment and fact recall. Participants were given a short survey that assigned them to (1) a random ordering of four font and pacing conditions and (2) a random ordering of four fictional stories. After reading each story, participants were asked to indicate their level of enjoyment with the reading task and to answer questions testing their recall of facts from the story. Results indicate that while the presentation of stories with a larger-than-normal font size has no consistently significant impact on reading enjoyment, applying forced-pacing to a reading experience significantly reduces reading enjoyment, constituting a main effect. Additionally, results suggest that neither font size nor pacing has a meaningful main effect on fact recall, and no significant interaction effects exist between font size and pacing on reading enjoyment or fact recall. These findings hold across several demographic and technological factors for English-speaking individuals — including age, sex, education level, and digital device screen resolution — indicating that observed effects are largely generalizable. They also offer valuable insights for the design of future digital reading experiences, optimized with user enjoyment in mind.

Introduction

For many people, reading is an integral part of daily life. While the act of reading might seem unexceptional, it actually involves a remarkable sequence of complex visual and neurological processes. The evolution of technology has driven more stories, books, and articles to digital formats, causing readers to increasingly rely on these processes to interpret text presented in a variety of stylistic formats, background colors, and mediums — including small and large font, light and dark color, and wide and narrow spacing.

Prior research shows that these stylistic differences can play a pivotal role in shaping an individual's reading experience. For example, font size and font type have been extensively studied for their effects on readability and reading performance measures. Hojjati et al. (2014)¹ and Wallace et al. (2022)² examined the impact of font characteristics on text readability and reading speed, and found that certain fonts or letter arrangements produced a more positive reading experience and allowed for quicker average reading completion times than others. Additionally, Katzir et al. (2013)³ and Rello et al. (2014)⁴ explored how font size affected reading comprehension across digital works, and found that larger fonts sometimes produced a positive impact on comprehension, depending on subject age.

¹Hojjati, Nafiseh and Muniandy, Balakrishnan. (2014). The Effects of Font Type and Spacing of Text for Online Readability and Performance. *Contemporary Educational Technology*. 5. 10.30935/cedtech/6122.

²Shaun Wallace, Zoya Bylinskii, Jonathan Dobres, Bernard Kerr, Sam Berlow, Rick Treitman, Nirmal Kumawat, Kathleen Arpin, Dave B. Miller, Jeff Huang, and Ben D. Sawyer. 2022. Towards Individuated Reading Experiences: Different Fonts Increase Reading Speed for Different Individuals. *ACM Trans. Comput.-Hum. Interact.* 29, 4, Article 38 (August 2022), 56 pages. <https://doi.org/10.1145/3502222>.

³Katzir T, Hershko S, Halamish V. The effect of font size on reading comprehension on second and fifth grade children: bigger is not always better. *PLoS One*. 2013;8(9):e74061. Published 2013 Sep 19. doi:10.1371/journal.pone.0074061.

⁴Rello, Luz and Pielot, Martin and Marcos, Mari-Carmen. (2016). Make It Big!: The Effect of Font Size and Line Spacing on Online Readability. 3637-3648. 10.1145/2858036.2858204.

We aim to build on this research and simultaneously address a current gap in the literature by exploring how different digital reading formats impact the way someone feels about a reading task. We do this by investigating the effects of various font sizes and pacing conditions — that is, how quickly one is moved (or paced) through a reading task — on (A) a given individual’s self-assessment of reading enjoyment (i.e., how much do they find a reading experience rewarding) and (B) that same individual’s success with recalling basic facts from a reading task. In this way, our study offers a unique perspective on reading behaviors and outcomes by placing the focus more on *experiential* measures of reading rather than *performative* measures of reading (e.g., reading speed and deeper comprehension).

Research Question, Theory, and Hypotheses

We summarize our key research question this way: ***Does font size, pacing, or the combination of both font size and pacing impact reading enjoyment and/or fact recall following a reading task?*** More specifically, we seek to explore whether or not a causal link exists between (1) individuals’ self-reported enjoyment and/or fact recall following a reading task and (2) the way that reading task is presented to them — with large font or default-sized font, and with forced-pacing or self-pacing.

We operate with the following null hypothesis and alternative hypothesis in mind:

Null Hypothesis (H_0): There is no significant interaction between font size and pacing and no main effects of font size or pacing on either reading enjoyment or fact recall.

Alternative Hypothesis (H_A): There are significant main effects of font size and/or pacing on reading enjoyment or fact recall, and a significant interaction between font size and pacing on reading enjoyment or fact recall.

While we are hesitant to apply *a priori* directional assumptions to H_A , we ultimately theorize that a larger font should yield higher reading enjoyment and lower fact recall than a smaller font. We suspect this is because larger fonts increase readers’ attention to content, decrease cognitive strain, and generate the perception of quicker progression through a task, which could both (A) tap into the brain’s reward system and produce feelings of accomplishment and happiness, and (B) make it easier to gloss over words and miss certain details in the text. We also theorize that the introduction of forced-pacing within a reading experience should yield lower reading enjoyment but higher fact recall. We suspect this is because forced-pacing has the potential to overwhelm some readers and reduce their sense of “control” or autonomy (i.e., the ability to go at their own pace) while reading — leading to an overall worse reading experience — yet also encourage increased focus on the details of the task to keep up with the speed in which a reading is presented.

Experiment Design and Treatment

To test our hypotheses, we conducted a within-subjects, survey-based experiment designed as a 2x2x4 factorial. The independent variables were font and pacing (termed “scrolling” within our experimental data and analysis), and four unique stories were randomly paired with each treatment condition to evaluate impacts on reading enjoyment (see the *Appendix* for copies of these stories). The stories were of similar length and difficulty, and were generated by ChatGPT.⁵ There were three different treatment conditions and one control condition. A large font condition presented a randomly-selected story in 36pt font one sentence at a time, requiring participants to click the “next” button in the survey to proceed after each sentence. A pacing condition was accomplished using JavaScript code that automatically scrolled through lines of 12pt text from a randomly-selected story at a specified word-per-minute (WPM) rate.⁶ An interaction condition combined both large font (36pt) and pacing conditions into a single experience for a randomly-selected story.⁷ Finally, a control condition presented a randomly-selected story on a single page with 12pt (default) font and

⁵OpenAI. (2023). ChatGPT 3.5 by OpenAI. Retrieved from <https://www.openai.com/research/chatgpt>.

⁶The exact WPM used was 280, which is slightly higher than the speed at which most individuals read.

⁷The “experience” of scrolling differed in the interaction condition depending on how much text was presented on the screen in a single block. Shorter blocks of text scrolled more quickly, while longer blocks of text scrolled more slowly.

no pacing, allowing participants to read the story at their own leisure. We employed the Qualtrics survey platform to design our full study experience and each reading task.⁸

Methods and Procedures

Prior to experiment launch, we conducted three pilot studies to refine our survey design and enhance user experience. All pilot studies and the final experiment were administered on Qualtrics. Participants for the final pilot and official experiment were recruited and compensated through Prolific. Study participants were required to be English-speaking residents of the US or UK, and they could participate in the experiment by taking the survey on a mobile phone, tablet, or computer. A commit-to-complete question was posed before any treatment administration to discourage attrition, and participant consent was also obtained.

In the survey, participants were first asked to rate their level of agreement with the statement “Reading is rewarding to me” on a 7-point Likert scale. This served as a baseline measurement for reading enjoyment. They were then presented with each of our four experimental conditions and readings. After each condition, participants were asked to rate their level of agreement with the statement “*This experience* of reading was rewarding to me” on the same 7-point Likert scale and were reminded of their baseline response. They were also asked to answer three multiple-choice questions about the story they just read to gauge fact recall. Finally, to minimize treatment persistence effects, participants were given a distractor task — which involved clicking and dragging words into appropriate bins/categories (e.g., “Paris” into “places”, “violin” into “musical instruments”, “tennis” into “sports”, etc.) — before proceeding to the next reading.

After completing all four readings, participants were asked two optional free-form/write-in questions: (1) “What factors influenced your response each time you read the statement: Reading is rewarding to me?” and (2) “Please provide any thoughts about the reading experiences presented today.” Participant comments on these two questions provided additional insights into the key factors contributing to their reading enjoyment scores and their overall study experience (as discussed briefly in the *Additional Insights* section).

Additional information about each participant was captured using metadata, Prolific-provided data, and pre-treatment study-specific questions. Primary variables of interest (model covariates) included device type, screen resolution, demographics, and whether participants used glasses or contacts for vision correction. reCAPTCHA was employed to deter bots and identify suspicious survey entries. Notably, a design oversight permitted 10 participants to take the survey multiple times. For these 10 individuals, data from only their first survey attempt was used within our analysis.

Our final participant sample size was $N = 195$. Figure 1 depicts how this value (N) was affected by factors like pre-treatment participant departure, non-consent, repeat survey-takers, and pre-launch test records (as discussed further in the *Data Collection* section).

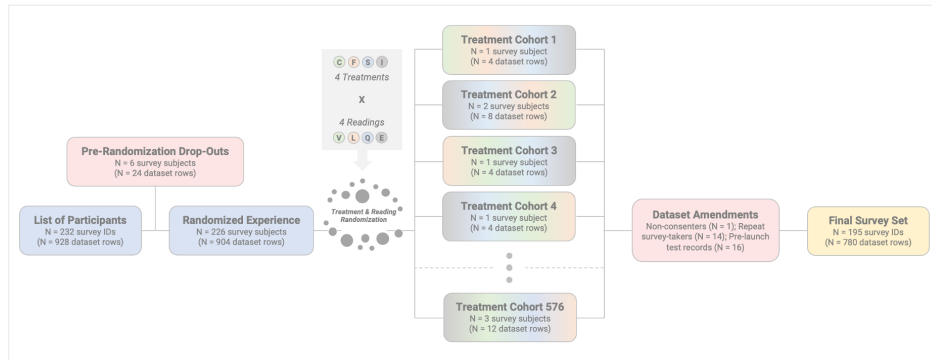


Figure 1: Experiment flow (CONSORT) diagram depicting study participants and sample size (N)

⁸The data for this paper was generated using Qualtrics software, Version 10-11/2023 of Qualtrics. Copyright © 2023 Qualtrics. Qualtrics and all other Qualtrics product or service names are registered trademarks or trademarks of Qualtrics, Provo, UT, USA. <https://www.qualtrics.com>.

Randomization and Outcome Measures

A total of 22 participants (approximately 11.3% of our study sample) attrited either during or after receiving at least one treatment. All other study participants experienced all four treatment conditions and read all four stories. We employed multiple randomizers in the Qualtrics survey flow to (A) assign participants to one of the 24 possible treatment-story combination blocks, and (B) shuffle the order of these combinations to generate 24 different arrangements per block. This yielded 576 equally likely unique treatment experiences/randomizations. Figure 2 displays a block of 24 possible treatment-story combinations, utilizing color to depict stories and subscripts to depict treatment conditions. For space and brevity, all 576 total unique arrangements of each treatment-story combination are not shown.

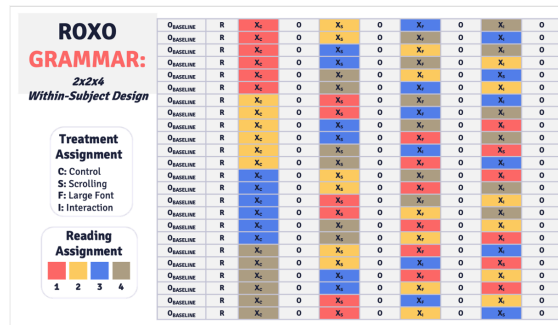


Figure 2: ROXO diagram depicting experimental design and within-subjects comparisons

Our primary outcome variable was reading enjoyment (termed “enjoyment” throughout our experimental data and analysis), measured on a discrete 7-point Likert scale ranging from “strongly disagree” to “strongly agree.” On this scale, a value of 1 (“strongly disagree”) signaled lowest enjoyment, and a value of 7 (“strongly agree”) signaled highest enjoyment.

Our secondary outcome variable was fact recall (termed “recall” throughout our experimental data and analysis). Participants were awarded 1 point for each correctly-answered recall question following a story. Each story had a maximum recall score of 3 points, and participants who answered no questions correctly received 0 points.

Power Calculation

We conducted a power analysis to estimate experimental power prior to analyzing study results. For this analysis, we created synthetic data that mimicked our final experiment design using one control and three treatment conditions, and we focused exclusively on reading enjoyment rather than recall.

In order to account for inherent differences between study participants (given our within-subjects experimental design), we created a variable called **offset** for power modeling, which corresponds to the baseline reading enjoyment value for each participant. We computed our final outcome variable (“enjoyment”) as a sum of a participant’s offset value and an assumed treatment effect, and utilized randomization inference to forecast various power values with different participant sample sizes. For simplicity, we assumed a modest treatment effect size of +0.25 (i.e., increasing enjoyment on average by 0.25 units on the 7-point Likert scale) for all treatment conditions. Under this assumption, we observed that our experiment would be well-powered even at low sample sizes (i.e., between 75-100 participants), as shown in Figure 3.

Experimental Data

Data Collection

Our final experimental dataset was comprised of (A) answers provided by study participants in our Qualtrics survey and (B) demographic/additional variables of interest (model covariates) for those participants provided

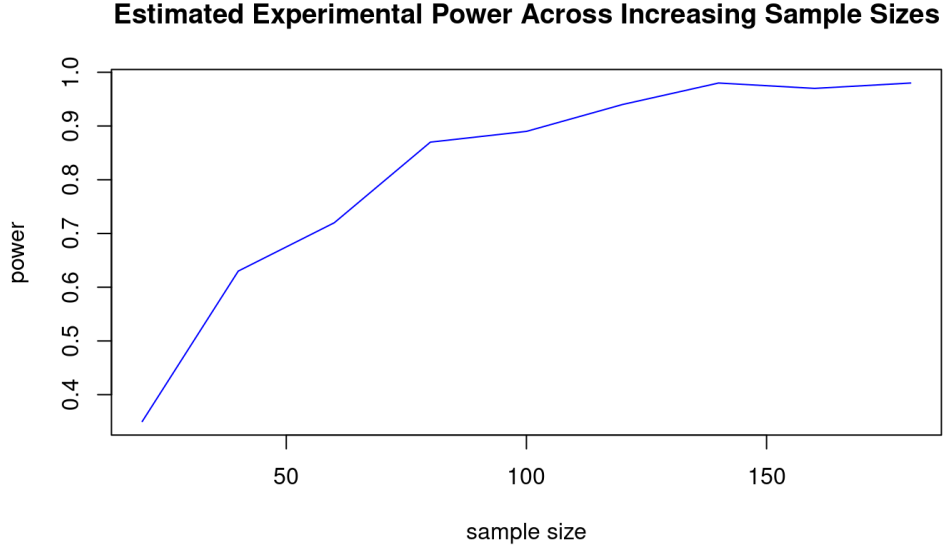


Figure 3: A pre-experiment power analysis chart showing estimated power under different sample size assumptions and an average treatment effect assumption of $+0.25$

by Prolific. A total of 232 survey responses were collected as part of the experiment. Of these 232 responses, 16 were from testing conducted prior to experiment launch and were excluded from the final dataset. Additionally, 1 survey record was associated with a participant who declined consent after clicking on the experiment survey link, 6 records were associated with participants who exited the survey before being randomized into our treatment conditions, and 14 records were associated with 10 unique individuals who took the survey either two or three times. These additional 21 records were excluded from the final dataset, and for all 10 repeat-survey takers, only the first survey attempt was included in the final dataset. This filtering process yielded a final dataset containing 195 unique survey responses. Note that because we employed four treatment conditions for each participant, our sample size of $N = 195$ translates to 780 rows ($780 = 195 \times 4$) in our final dataset.

Figure 4 shows the distributions of participants across our outcomes of interest — reading enjoyment and fact recall — categorized by whether each outcome was measured in the control, font, pacing (scrolling), or interaction condition. Both distributions are left-skewed, with enjoyment concentrated in the 5-7 region, and recall in the 2-3 region.

Incomplete Data

Exploratory data analysis (EDA) revealed that some of the demographic/additional variables of interest provided by Prolific were missing from participant records. For example, there were 17 participants for which the **Sex** variable was missing from the data, and a data value of “*Unspecified*” was used to populate these missing values. There were also 18 participants for which the **Age** variable was missing, and a data value of 99 was used to populate these missing values as an encoding for “*unknown*” (prior to this update, the previous maximum age in the data set was 91).

Analysis also revealed 22 participants who attrited from the experiment, resulting in a total of 65 data rows which had missing enjoyment and recall measurements. As attrition can significantly impact the estimate of treatment effects, two additional datasets were prepared for extreme value bounds (EVB) analysis: a “lower bound” dataset, and an “upper bound” dataset. In the lower bound dataset, a value of 1 (the minimum possible value from the Likert scale for gauging reading enjoyment) was imputed to any missing enjoyment records, and a value of 0 (the minimum possible value for story recall questions) was imputed to any missing

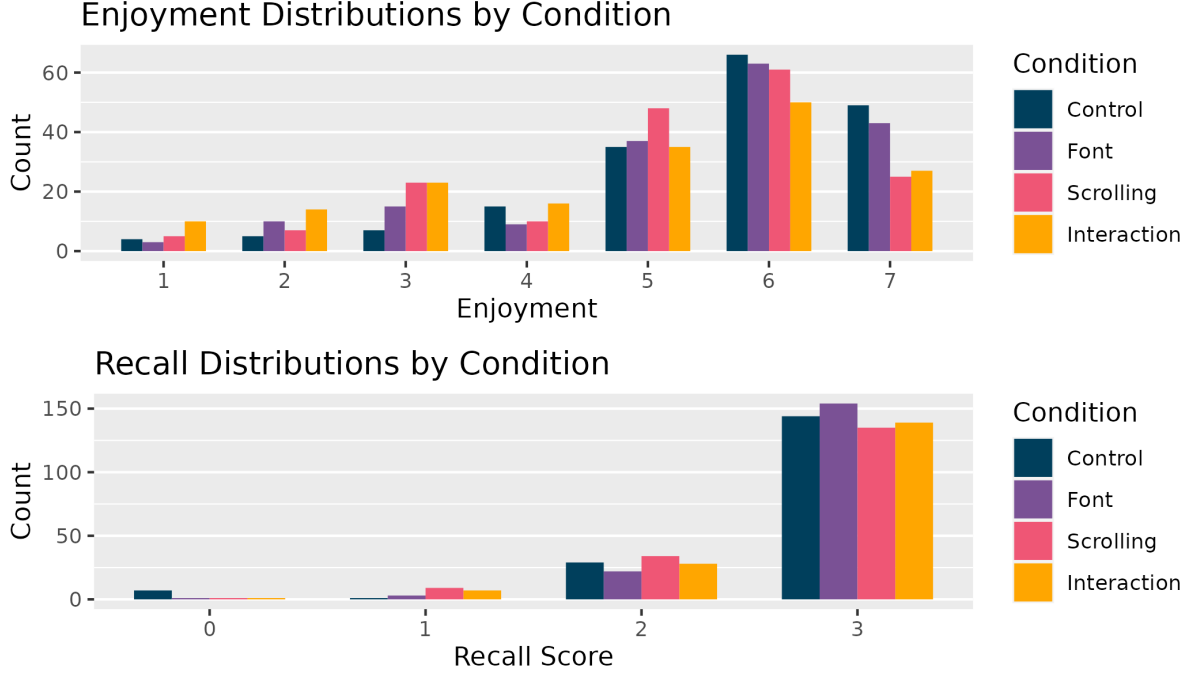


Figure 4: Distribution of participants by experimental outcome measures: reading enjoyment (top) and fact recall (bottom)

recall records. For the upper bound dataset, a value of 7 (the maximum possible value from the Likert scale for gauging reading enjoyment) was imputed to any missing enjoyment records, and a value of 3 (the maximum possible value for story recall questions) was imputed to any missing recall records. The resulting impacts on our overall (non-treatment specific) outcome distributions are displayed in Figure 5.

Results and Models

To understand the effects of treatment on outcomes, we evaluated a progression of modeling approaches for each outcome of interest. The first approach is a difference in means model. The second approach increases complexity by accounting for treatment-specific effects. The third approach builds on the second by including round effects. In general for estimating both outcome variables — reading enjoyment (**enjoyment**) and fact recall (**recall**) — we fit regressions of the form,

$$(1) \hat{y} = \beta_0 + \beta_1 \cdot \text{baseline} + \mathbf{Z}\gamma,$$

$$(2) \hat{y} = \beta_0 + \beta_1 \cdot \text{baseline} + \beta_2 \cdot \text{font} + \beta_3 \cdot \text{scrolling} + \beta_4 \cdot (\text{font} * \text{scrolling}) + \mathbf{Z}\gamma, \text{ and}$$

$$(3) \hat{y} = \beta_0 + \beta_1 \cdot \text{baseline} + \beta_2 \cdot \text{font} + \beta_3 \cdot \text{scrolling} + \beta_4 \cdot (\text{font} * \text{scrolling}) + \beta_5 \cdot r_2 + \beta_6 \cdot r_3 + \beta_7 \cdot r_4 + \mathbf{Z}\gamma,$$

where \hat{y} is a modeled estimate for either **enjoyment** or **recall**, β_0 is a constant, β_1 is the change to the estimate (“marginal change”) driven by each unit of increase in an individual’s **baseline** level of reading enjoyment, β_2 is the marginal change driven by the presence of the **font** treatment condition, β_3 is the marginal change driven by the presence of the **scrolling** (or pacing) treatment condition, β_4 is the *additional* marginal change *beyond* the marginal changes driven by the **font** and **scrolling** coefficients of experiencing both the **font** and **scrolling** treatments concurrently, β_5 is the marginal change resulting from r_2 (round 2) of treatment, β_6 is the marginal change resulting from r_3 (round 3) of treatment, β_7 is the marginal change resulting from r_4 (round 4) of treatment, \mathbf{Z} is a row vector of additional covariates, and γ is a column vector of coefficients.

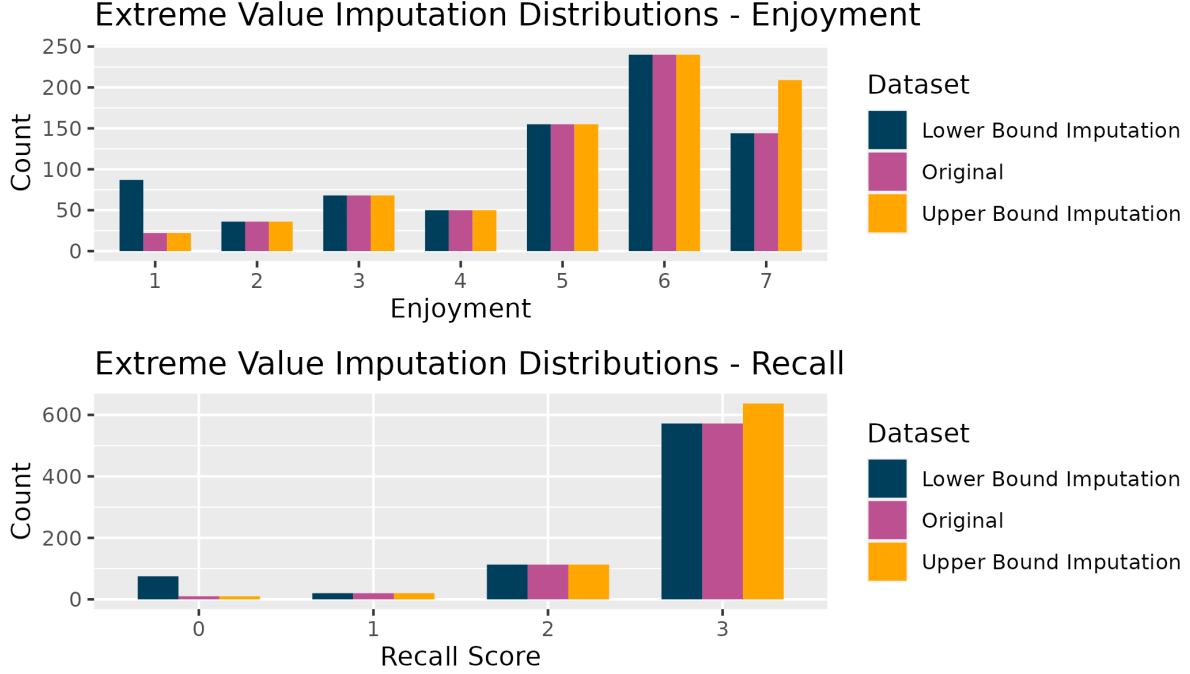


Figure 5: Distribution of participants by experimental outcome measures with EVB applied: reading enjoyment (top) and fact recall (bottom)

Within each of these approaches, we evaluated four distinct models. The first model (“Base Model”) includes our treatment coefficients of interest but no covariates. The second model (“Base Model + Covariates”) adds individual fixed effects. These effects include **mobile device**⁹, **education level**, **glasses/contacts group**¹⁰, **screen resolution**, **age**, and **sex** indicators. The third and fourth models (“Lower Bound Imputation” and “Upper Bound Imputation”) estimate treatment effects with lower bounds or upper bounds imputed/applied to missing data, respectively. These models also include individual fixed effects and serve to bound the true ATE (average treatment effect) for each treatment condition given experimental attrition.

In the following model summary tables, statistical significance is calculated using Benjamini-Hochberg (BH) corrected p-values. Additionally, clustered standard errors — where clustering occurs on participants’ Prolific IDs — are reported in parentheses.

Reading Enjoyment

Table 1 shows the results of the difference in means models for the enjoyment outcome. Immediately apparent is the high statistical significance of the **Treatment** coefficient across these models. Notably, and contrary to our initial pre-experiment theory, the **Treatment** coefficient is negative for all of the models, meaning treatment conditions on average actually *decrease* enjoyment.

The **Baseline Enjoyment** coefficient is also statistically significant across these models. This is an indication that reading enjoyment is a persistent trait and is predictive of enjoyment, regardless of treatment condition.

Table 2 builds on the previous series of models by estimating treatment-specific effects. This is accomplished by utilizing indicator variables for when a participant is in the font treatment, the pacing treatment, and the interaction (combination of font and pacing) treatment. Note that when a participant is in the interaction

⁹Mobile device indicates whether or not the participant took the survey on a mobile device, like a cell phone.

¹⁰Glasses/contacts groups include the following categorizations: (1) Participant does not require corrective lenses, (2) Participant requires corrective lenses for reading but was not wearing them at the time of experiment, (3) Participant usually requires corrective lenses and was wearing them for the experiment, and (4) Participant preferred not to answer.

Table 1: Difference in Means Models (Enjoyment Outcome)

	<i>Dependent variable:</i>			
	Base Model	Base Model + Covariates	Enjoyment (Likert Scale from 1-7) Lower Bound Imputation	Upper Bound Imputation
	(1)	(2)	(3)	(4)
Treatment	-0.501*** (0.094)	-0.452*** (0.103)	-0.523*** (0.114)	-0.431*** (0.101)
Baseline Enjoyment	0.546*** (0.082)	0.551*** (0.092)	0.409*** (0.107)	0.431*** (0.132)
Constant	2.319*** (0.484)	0.317 (1.173)	0.435 (1.380)	1.153 (1.330)
Individual Fixed Effects	No	Yes	Yes	Yes
Observations	715	715	780	780
R ²	0.153	0.461	0.499	0.424
Adjusted R ²	0.151	0.352	0.408	0.320
Residual Std. Error	1.457 (df = 712)	1.272 (df = 594)	1.469 (df = 659)	1.314 (df = 659)
F Statistic	64.501*** (df = 2; 712)	4.239*** (df = 120; 594)	5.472*** (df = 120; 659)	4.050*** (df = 120; 659)

Note:

* p<0.1; ** p<0.05; *** p<0.01
 Clustered standard errors in parentheses
 Statistical significance calculated using Benjamini-Hochberg-corrected p-values
 Individual Fixed Effects include: Mobile Device, Education Level, Glasses/Contacts Group, Screen Resolution, Age, and Sex

treatment, the **Font**, **Scrolling**, and **Font:Scrolling** indicators are all active and their corresponding coefficients sum to give the ATE of the interaction treatment.

These models, like the previous ones, show that **Baseline Enjoyment** is a strong predictor of enjoyment regardless of treatment condition. Additionally, the negative sign on all treatment coefficients indicates that treatment conditions tend to reduce reading enjoyment. The pacing (scrolling) treatment shows a larger and more statistically significant effect on decreasing reading enjoyment than the font treatment. This effect remains statistically significant for the “Upper Bound Imputation” model. That is, even in the scenario where reading enjoyment is assumed to be *maximized* for participants who attrited, there is evidence to suggest that forced-pacing still decreases reading enjoyment.

Table 2: Treatment-Specific Models (Enjoyment Outcome)

	<i>Dependent variable:</i>			
	treat Model	treat Model + Covariates	Enjoyment (Likert Scale from 1-7) Lower Bound Imputation	Upper Bound Imputation
	(1)	(2)	(3)	(4)
Font	-0.209 (0.110)	-0.134 (0.119)	-0.205 (0.136)	-0.174 (0.117)
Scrolling	-0.493*** (0.102)	-0.460*** (0.112)	-0.503*** (0.121)	-0.441*** (0.111)
Baseline Enjoyment	0.545*** (0.083)	0.550*** (0.092)	0.409*** (0.107)	0.431*** (0.132)
Font:Scrolling	-0.107 (0.140)	-0.173 (0.150)	-0.154 (0.155)	-0.062 (0.150)
Constant	2.324*** (0.484)	0.323 (1.174)	0.435 (1.383)	1.153 (1.332)
Individual Fixed Effects	No	Yes	Yes	Yes
Observations	715	715	780	780
R ²	0.171	0.481	0.514	0.437
Adjusted R ²	0.167	0.374	0.424	0.332
Residual Std. Error	1.443 (df = 710)	1.251 (df = 592)	1.449 (df = 657)	1.301 (df = 657)
F Statistic	36.689*** (df = 4; 710)	4.498*** (df = 122; 592)	5.694*** (df = 122; 657)	4.179*** (df = 122; 657)

Note:

* p<0.1; ** p<0.05; *** p<0.01
 Clustered standard errors in parentheses
 Statistical significance calculated using Benjamini-Hochberg-corrected p-values
 Individual Fixed Effects include: Mobile Device, Education Level, Glasses/Contacts Group, Screen Resolution, Age, and Sex

Table 3 builds on the previous model by including round effects. This is accomplished by using indicator variables for **Second Round**, **Third Round**, and **Fourth Round** (implying that the **Constant** incorporates the effect of the **First Round**).

Due to the randomization design of the experiment, the treatment-specific coefficient estimates change only marginally with the inclusion of the round effects coefficients. The **Second Round** coefficients show a negative effect on enjoyment, but none of these estimates are significantly different from no effect on enjoyment. However, both the **Third Round** and **Fourth Round** coefficients have a statistically significant negative effect

on reading enjoyment. Potential explanations for this result include participant fatigue toward the end of the experiment or persistent effects from previous treatment conditions.

Table 3: Round Effects Models (Enjoyment Outcome)

	Dependent variable:			
	Base Model	Base Model + Covariates	Enjoyment (Likert Scale from 1-7)	
			Lower Bound Imputation	Upper Bound Imputation
	(1)	(2)	(3)	(4)
Font	-0.228 (0.108)	-0.155 (0.118)	-0.236 (0.131)	-0.184 (0.116)
Scrolling	-0.500*** (0.102)	-0.469*** (0.111)	-0.511*** (0.116)	-0.445*** (0.111)
Baseline Enjoyment	0.546*** (0.083)	0.551*** (0.092)	0.409*** (0.107)	0.431*** (0.133)
Second Round	-0.195 (0.115)	-0.188 (0.128)	-0.407 (0.133)	-0.092 (0.123)
Third Round	-0.335* (0.120)	-0.353* (0.131)	-0.629* (0.138)	-0.159* (0.132)
Fourth Round	-0.366** (0.112)	-0.382** (0.122)	-0.688** (0.136)	-0.194** (0.124)
Font:Scrolling	-0.096 (0.140)	-0.162 (0.151)	-0.138 (0.152)	-0.054 (0.150)
Constant	2.551*** (0.484)	0.556 (1.168)	0.882 (1.375)	1.269 (1.323)
Individual Fixed Effects	No	Yes	Yes	Yes
Observations	715	715	780	780
R ²	0.180	0.490	0.534	0.439
Adjusted R ²	0.172	0.382	0.445	0.332
Residual Std. Error	1.439 (df = 707)	1.243 (df = 589)	1.422 (df = 654)	1.302 (df = 654)
F Statistic	22.144*** (df = 7; 707)	4.532*** (df = 125; 589)	5.994*** (df = 125; 654)	4.096*** (df = 125; 654)

Note:

*p<0.1; **p<0.05; ***p<0.01
 Clustered standard errors in parentheses
 Statistical significance calculated using Benjamini-Hochberg-corrected p-values
 Individual Fixed Effects include: Mobile Device, Education Level, Glasses/Contacts Group, Screen Resolution, Age, and Sex

Fact Recall

We applied the same model progression to evaluate fact recall. Table 4 shows the results of the difference in means models. Here, the **Constant** term is highly statistically significant across all models and approximates the maximum recall score of 3. Potential explanations for this outcome include simple readings, simple recall questions, or high intelligence across experimental participants. On average, our experimental treatments had no impact on recall.

Table 4: Difference in Means Models (Recall Outcome)

	Dependent variable:			
	Base Model	Base Model + Covariates	Recall (Scored from 0-3)	
			Lower Bound Imputation	Upper Bound Imputation
	(1)	(2)	(3)	(4)
Treatment	0.042 (0.049)	0.014 (0.047)	-0.003 (0.049)	0.043 (0.051)
Baseline Enjoyment	0.011 (0.022)	0.020 (0.036)	0.005 (0.050)	0.016 (0.027)
Constant	2.646*** (0.151)	2.916*** (0.434)	2.409*** (0.586)	2.768*** (0.394)
Individual Fixed Effects	No	Yes	Yes	Yes
Observations	715	715	780	780
R ²	0.001	0.291	0.530	0.233
Adjusted R ²	-0.001	0.147	0.445	0.094
Residual Std. Error	0.576 (df = 712)	0.531 (df = 594)	0.699 (df = 659)	0.528 (df = 659)
F Statistic	0.513 (df = 2; 712)	2.029*** (df = 120; 594)	6.202*** (df = 120; 659)	1.672*** (df = 120; 659)

Note:

*p<0.1; **p<0.05; ***p<0.01
 Clustered standard errors in parentheses
 Statistical significance calculated using Benjamini-Hochberg-corrected p-values
 Individual Fixed Effects include: Mobile Device, Education Level, Glasses/Contacts Group, Screen Resolution, Age, and Sex

Table 5 includes treatment-specific effects for the models estimating recall. The **Font** coefficients are marginally greater than zero and the **Scrolling** coefficients are marginally less than zero. The lack of statistical significance and the marginal size of the effects preclude practical significance for these treatment conditions.

Table 5: Treatment-Specific Models (Recall Outcome)

	<i>Dependent variable:</i>			
			Recall (Scored from 0-3)	
	treat Model	treat Model + Covariates	Lower Bound Imputation	Upper Bound Imputation
	(1)	(2)	(3)	(4)
Font	0.115 (0.053)	0.088 (0.050)	0.092 (0.056)	0.108 (0.055)
Scrolling	-0.020 (0.060)	-0.044 (0.062)	-0.046 (0.065)	-0.015 (0.062)
Baseline Enjoyment	0.011 (0.022)	0.020 (0.036)	0.005 (0.050)	0.016 (0.027)
Font:Scrolling	-0.065 (0.076)	-0.044 (0.080)	-0.103 (0.084)	-0.056 (0.078)
Constant	2.648*** (0.151)	2.920*** (0.433)	2.409*** (0.587)	2.768*** (0.394)
Individual Fixed Effects	No	Yes	Yes	Yes
Observations	715	715	780	780
R ²	0.008	0.298	0.534	0.240
Adjusted R ²	0.003	0.153	0.448	0.098
Residual Std. Error	0.574 (df = 710)	0.529 (df = 592)	0.697 (df = 657)	0.527 (df = 657)
F Statistic	1.520 (df = 4; 710)	2.056*** (df = 122; 592)	6.179*** (df = 122; 657)	1.697*** (df = 122; 657)

Note:

* p<0.1; ** p<0.05; *** p<0.01
 Clustered standard errors in parentheses
 Statistical significance calculated using Benjamini-Hochberg-corrected p-values
 Individual Fixed Effects include: Mobile Device, Education Level, Glasses/Contacts Group, Screen Resolution, Age, and Sex

Table 6 adds round effects to these models. Similar to the previous recall models, the **Constant** coefficient is most statistically significant and predictive. Meanwhile, treatment or round effects are not statistically significantly different from having no effect.

Table 6: Round Effects Models (Recall Outcome)

	<i>Dependent variable:</i>			
			Recall (Scored from 0-3)	
	Base Model	Base Model + Covariates	Lower Bound Imputation	Upper Bound Imputation
	(1)	(2)	(3)	(4)
Font	0.113 (0.053)	0.085 (0.050)	0.081 (0.054)	0.107 (0.056)
Scrolling	-0.021 (0.060)	-0.046 (0.061)	-0.049 (0.065)	-0.016 (0.062)
Baseline Enjoyment	0.011 (0.022)	0.020 (0.036)	0.005 (0.050)	0.016 (0.027)
Second Round	0.012 (0.055)	-0.008 (0.057)	-0.132 (0.063)	0.025 (0.057)
Third Round	-0.029 (0.056)	-0.067 (0.056)	-0.240 (0.065)	-0.005 (0.059)
Fourth Round	0.016 (0.055)	-0.024 (0.055)	-0.212 (0.067)	0.035 (0.058)
Font:Scrolling	-0.065 (0.077)	-0.043 (0.081)	-0.098 (0.083)	-0.056 (0.078)
Constant	2.649*** (0.159)	2.945*** (0.442)	2.561*** (0.589)	2.755*** (0.402)
Individual Fixed Effects	No	Yes	Yes	Yes
Observations	715	715	780	780
R ²	0.009	0.300	0.544	0.240
Adjusted R ²	-0.0004	0.151	0.457	0.095
Residual Std. Error	0.575 (df = 707)	0.530 (df = 589)	0.691 (df = 654)	0.528 (df = 654)
F Statistic	0.962 (df = 7; 707)	2.015*** (df = 125; 589)	6.247*** (df = 125; 654)	1.657*** (df = 125; 654)

Note:

* p<0.1; ** p<0.05; *** p<0.01
 Clustered standard errors in parentheses
 Statistical significance calculated using Benjamini-Hochberg-corrected p-values
 Individual Fixed Effects include: Mobile Device, Education Level, Glasses/Contacts Group, Screen Resolution, Age, and Sex

Additional Qualitative Insights

For additional insights on the drivers of our experimental results, we briefly analyzed the comments left by participants within the optional free-form questions at the end of our survey. A total of 169 unique participants left comments under one or both of these questions. We applied basic NLP preprocessing techniques to (A) construct two separate vocabularies corresponding to the comments data for the two free-form survey questions — removing punctuation, numbers, and any English “stopwords” (as well as the

unusually common words “read” and “reading” in participant responses based on our survey subject matter) along the way — and (B) generate word clouds based on term frequencies within each vocabulary.¹¹ A separate word cloud was generated for each of our two free-form survey questions.

The resulting word clouds are shown in Figure 6. These visualizations allow us to quickly see which words most frequently appeared within participant comments, with word size denoting word frequency (i.e., larger size indicates higher frequency, smaller size indicates lower frequency). Notably, words like “scrolling”, “scrolled”, “pace”, and “paced” appear in moderate to large size across both visuals, suggesting that our pacing treatment impacted participants’ reading experiences to such a degree that it was frequently mentioned in survey comments. We also see similar pacing-related words like “fast”, “faster”, “speed”, “quickly”, “moved”, “moving”, and “control” appear in one or both visuals, which validates the presence of a main effect with forced-pacing. Additionally, both word clouds clearly feature words like “enjoy”, “enjoyed”, “text”, “presented”, and “experience”, which supplies evidence that participants actually engaged with our intended experimental concepts and/or story design characteristics. Conversely, words like “font”, “size”, “big”, and “small” appear to be either absent from or depicted in smaller-sized font across one or both visuals, suggesting that font size differences across the reading experiences didn’t register as consistently impactful to participants’ experiences (or, at least, not nearly as much as pacing differences did).

These insights provide an additional layer of *qualitative* support for the conclusions drawn from our *quantitative* results. They corroborate the findings from our models and indicate that participant reading experiences were indeed *affected* by our treatment conditions. With additional time, we would take this analysis further by applying NLP sentiment classification techniques to investigate comment positivity/negativity (both alone and interacted with treatment condition and reading enjoyment data), which could help us determine if any of our treatments — or any changes observed in enjoyment across rounds — correlated with particular participant sentiments and feelings.

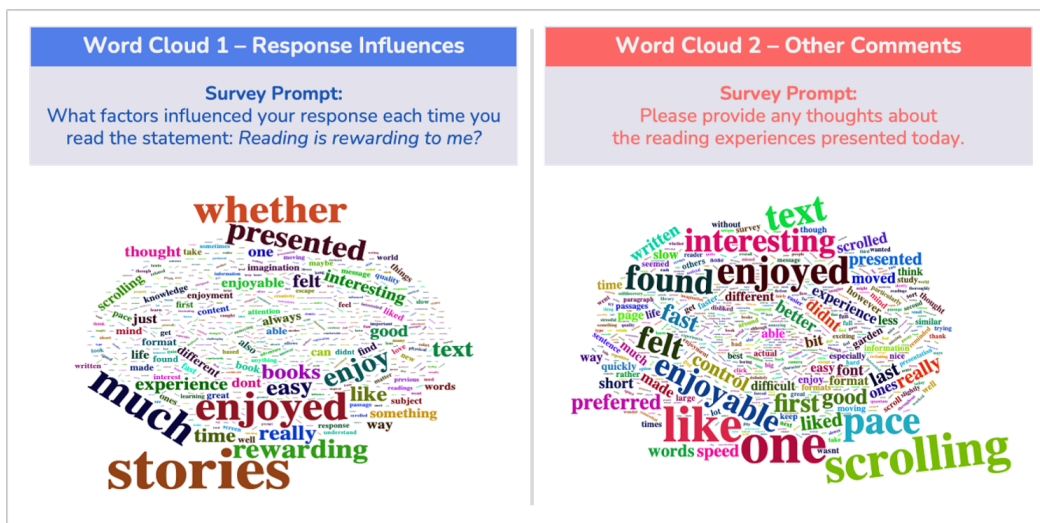


Figure 6: Word Cloud visualizations of most frequent words in participant survey comments: response influences (left) and other comments (right); note: common English stopwords are excluded, along with the words "read" and "reading"

Conclusion

Our experimental results lead to separate conclusions for our outcome measures of interest. First, with respect to *reading enjoyment*, we ultimately reject our null hypothesis and the directionality of effects posed in our alternative hypothesis. We find that, as a whole, some of our treatments have statistically significant

¹¹For a record of all 174 "stopwords" in the English language list, see the R documentation found here: <https://rdrr.io/rforge/tm/man/stopwords.html>.

negative main effects on enjoyment at a level that is also of practical significance. Specifically, we conclude that forced-pacing (scrolling) has the largest impact on enjoyment at statistically significant levels, even after applying EVB analysis to account for participant attrition. We also find that statistically significant negative round effects exist for the third and fourth rounds of treatment, which is suggestive of participant fatigue or persistence effects from previous (i.e., first or second round) treatment conditions.

Second, with respect to *fact recall*, we fail to reject our null hypothesis. We find no statistically significant or practically significant main effects, interaction effects, or round effects on recall for our treatment conditions.

To build on our findings, future research could test whether our observed effects manifest across additional reading mediums — such as paper textbooks and audio books — and/or additional variations of font size, font color, stylistic/formatting details, or forced-pacing conditions (e.g., larger font, *slower* pacing). Additionally, we wonder if the incorporation of a “progress bar” (a visual/feature that explicitly indicates advancement toward a goal) within any of the reading experiences we presented might lead to different results, especially when testing larger font sizes. Finally, we suggest future studies explore the link between forced-pacing and reading comprehension, moving beyond the basic measurement of fact recall we employed within our research.

A key implication of our findings is this: allowing readers the autonomy to choose the pace at which they read and giving them the ability to adjust the font size of their reading material plays an important role in the reading experience. In fact, when this “control” is stripped away, we see meaningful impacts to reading enjoyment. Practically, this result holds importance for educators, reading instructors, and companies that design digital reading tools or products for the general population. In light of our findings, each of these groups can be increasingly cognizant of how the design features within a created or administered reading experience affects reader enjoyment — either for the better, or for the worse.

Appendix: ChatGPT-Generated Stories & Qualtrics Survey Link

Story A: The Last Voyage of the Space Explorer

Captain Adrian Rourke gazed through the large, transparent viewport of his spacecraft, the “Stardancer,” as it hurtled through the endless expanse of space. At the ripe age of 72, he was embarking on his final voyage, determined to unlock the mysteries of a cosmic anomaly known as the “Eternal Halo.”

The stars twinkled like a sea of diamonds, and the distant glow of the Eternal Halo beckoned him closer. Decades of exploration had led him to this very moment, and despite the physical toll his body had endured, his spirit burned with an unquenchable thirst for knowledge.

Adrian couldn’t help but feel the weight of his age and the realization that this might be his last journey. He thought about the years of wonder and discovery, the adventures and the solitude. A sense of nostalgia enveloped him as he considered the path he had chosen.

The ship’s computer, an AI named LARA, chimed in. “Captain Rourke, we are approaching the Eternal Halo’s event horizon. Scans indicate increased gravitational forces and intense radiation.”

Adrian nodded, his rugged face showing a hint of weariness. “Prepare for entry, LARA. This is it.”

As the “Stardancer” passed through the shimmering boundary of the anomaly, the laws of physics seemed to distort. Time became fluid, and space stretched and contracted. The past and future danced in an intricate waltz, and for a moment, Adrian felt as though he had glimpsed the very essence of the cosmos.

With each passing moment in the heart of the Eternal Halo, Adrian’s mind oscillated between the wonder of the unknown and the longing for the familiar. He couldn’t help but think of the countless nights he had spent gazing at the stars, yearning to understand their mysteries.

He couldn’t help but reflect on the life he had led. The countless planets he had visited, the discoveries he had made, and the sacrifices he had endured. He thought of his beloved wife, Elena, who had left him all those years ago, unable to bear the weight of his insatiable curiosity.

LARA’s voice broke his reverie. “Captain, we have reached the core of the Eternal Halo. Sensors indicate a profound source of energy. Should I initiate data collection?”

Adrian’s heart swelled with anticipation. “Yes, LARA, it’s time to uncover the secrets of the universe.”

As the “Stardancer” began its data collection process, an intense surge of energy pulsed through the ship. Adrian watched in awe as the console before him displayed readings that defied comprehension. The knowledge contained within this anomaly was beyond anything he had ever encountered.

But as he delved deeper into the cosmic enigma, he couldn’t help but feel a sense of closure. He had lived a life devoted to exploration, but it had come at a cost. The years spent away from his family, the sacrifices made in the name of discovery—all had led him to this moment.

With the data collection complete, Adrian knew that his journey was over. The “Stardancer” began its journey back to the familiar realms of space. He couldn’t help but smile as he thought of the generations of explorers who would benefit from the knowledge he had gathered.

As the spacecraft left the Eternal Halo behind, Captain Adrian Rourke felt a profound sense of fulfillment. He had lived a life of purpose and curiosity, and now, it was time to return to the people and places he had left behind.

For even in the vastness of space, there was a universal truth that he had rediscovered—exploration was not merely a quest for the unknown, but a journey toward understanding and connection, and that journey could lead one back home.

As the “Stardancer” neared Earth, a sense of peace washed over him. He looked forward to reuniting with his family, sharing the wisdom he had gained, and finally experiencing the quiet moments that he had sacrificed for the call of the cosmos. His final voyage had brought him full circle, from the depths of space to the warmth of home, reminding him that the greatest discoveries were often found in the simplest of connections.

Story B: The Library of Lost Stories

Eleanor had always been an unassuming librarian in the small town of Willowbrook. Her life was quiet, and her days were filled with the hushed whispers of the town's avid readers. She had a fondness for books, old and new, and her library was a sanctuary of stories waiting to be explored.

One chilly autumn evening, as Eleanor was shelving books in the back of the library, she discovered a hidden door tucked away behind a dusty bookshelf. Curiosity piqued, she turned the tarnished brass handle, and the door creaked open to reveal a dimly lit room.

Before her lay a vast and dusty library, unlike anything she had ever seen. The shelves were lined with leather-bound tomes, each one bearing a unique title: "The Unfinished Symphony," "The Vanishing Moon," "The Forgotten Promise," and countless others. These were not ordinary books; they were the Library of Lost Stories.

Eleanor's fingers trembled as she pulled one of the books from the shelf. The pages crackled with age, and as she opened it, words flowed across the pages, forming a story that had never been told. The words seemed to come to life, painting vivid scenes and compelling characters.

She spent hours reading, completely entranced by the stories that had been hidden from the world. Each book contained tales of love, adventure, and mystery, and Eleanor felt like she had discovered a hidden treasure trove of the human imagination.

As days turned into weeks, Eleanor dedicated herself to cataloging the books and reading each one cover to cover. She realized that these were stories that had been started but never finished, dreams that had been abandoned by their creators. Now, they had found their way to her.

One day, while reading a story about a brave young explorer searching for a lost city in the Amazon, Eleanor was struck by an idea. She would complete these stories, breathe life into them, and share them with the world. She would become the guardian of the Library of Lost Stories.

Her determination knew no bounds. She spent countless hours writing, editing, and revising the tales within the library. The characters that had been left in limbo now had their destinies fulfilled, and the stories were brought to completion.

With each story she finished, a sense of fulfillment washed over her. She felt like she was giving a voice to the forgotten characters, allowing their adventures to be told. And with each completed story, the shelves of the library seemed to shine a bit brighter.

Word of Eleanor's mission began to spread throughout Willowbrook. People from the town and neighboring areas came to read the newly finished stories and marvel at the extraordinary library. It became a place of inspiration and wonder, where the dreams of authors long gone were given new life.

One afternoon, a young woman named Emma approached Eleanor with a timid request. She had started writing a novel but had lost her inspiration and didn't know how to finish it. Eleanor, who had found her own inspiration in the Library of Lost Stories, welcomed Emma and introduced her to the hidden world of unfinished tales.

Together, they embarked on a journey to complete Emma's novel, and as they breathed life into her characters and story, a strong bond formed between them. Eleanor's mission had not only given new life to forgotten stories but also connected her with people in search of their own creative voices.

As the years passed, Eleanor's hair turned silver, but her spirit remained as youthful as ever. The Library of Lost Stories had grown, and its shelves held tales from every corner of the world. Eleanor's library had become a hub of creativity and connection, where the boundaries between readers, writers, and the stories themselves blurred.

And so, Eleanor, the unassuming librarian of Willowbrook, had transformed her quiet life into a vibrant tapestry of stories, where the lost had been found, and the forgotten had been remembered. In the Library of Lost Stories, words had found their way home, and in Eleanor's heart, a new chapter of her own story had begun.

Story C: The Quantum Photograph

In the heart of the bustling city of New York, where the relentless energy of life never seemed to fade, lived a scientist named Dr. Evelyn Reynolds. She had dedicated her life to the study of quantum physics, a field that both fascinated and challenged her intellect.

One day, while rummaging through a dusty thrift shop, Evelyn discovered an antique camera unlike any she had ever seen. The shopkeeper told her it was a “Quantum Camera” and that it had once belonged to a renowned scientist.

Intrigued, Evelyn purchased the camera and brought it back to her small, cluttered apartment. She studied the camera’s intricacies, noting its delicate brass dials and the peculiar glass lens. A handwritten note tucked into the camera’s leather case piqued her curiosity further. It read, “Capture the invisible.”

Evelyn, always eager for a scientific challenge, decided to experiment with the camera. She took it to her lab, set up a controlled environment, and captured her first photograph. To her amazement, the image revealed a strange, glowing aura surrounding an everyday object.

Word of Evelyn’s discovery spread throughout the scientific community, and she was soon invited to present her findings at a prestigious conference. The audience of brilliant minds marveled at her breakthroughs in capturing the unseen.

But Evelyn wasn’t satisfied. She realized that her camera could do more than just capture these invisible forces; it could provide a window into alternate realities. She had a vision of using the Quantum Camera to peer into parallel universes, to see the choices and paths not taken.

Late one night in her lab, Evelyn set up the camera for an audacious experiment. She aimed it at a photograph of herself from years ago, a pivotal moment when she had chosen one career path over another. With a click, the camera’s shutter snapped, and the image developed before her eyes.

To her astonishment, the photograph revealed an alternate version of herself, a scientist who had pursued a different path, living a life of adventure and discovery. It was as if the Quantum Camera had torn a hole in the fabric of reality and given her a glimpse into a parallel universe.

Evelyn was both thrilled and haunted by what she had witnessed. The possibilities were endless, and the temptation to explore these alternate realities grew stronger with each passing day. She became obsessed, spending sleepless nights perfecting her experiments.

As Evelyn delved deeper into the mysteries of the Quantum Camera, she found herself navigating a maze of branching timelines, each with its own set of choices and consequences. The camera had unlocked a Pandora’s box of endless possibilities, and Evelyn was consumed by the desire to explore them all.

One day, as she prepared to take another journey into the unknown, Evelyn noticed the note in the camera’s leather case once more. “Capture the invisible,” it read. The words took on a new meaning for her. She realized that the most precious and invisible thing of all was the life she had chosen, the path she had taken, and the reality she had created.

With a heavy heart, Evelyn made a decision. She turned off the Quantum Camera and locked it away in a vault, vowing never to use it again. She had come to understand that the pursuit of infinite possibilities had blinded her to the beauty of the reality she had chosen.

Evelyn returned to her life as a scientist in the bustling city of New York. She cherished each moment, each decision, and each consequence as a unique and precious part of her own reality. She had learned that it was not the endless choices that gave life meaning, but the choices themselves, and the journey that came with them.

Story D: The Enchanted Garden

Once upon a time, in a small cottage nestled at the edge of a dense forest, lived a young girl named Emily. Her family had recently moved to this tranquil place, hoping for a fresh start. As Emily explored her new surroundings, she discovered a hidden, overgrown garden tucked away behind the cottage.

The garden was unlike any she had seen before. Vibrant flowers of every color imaginable, towering trees with leaves that seemed to whisper secrets, and a pond with water so clear it mirrored the sky. The air was filled with the enchanting scent of blossoms, and the garden itself was overgrown with ivy and wildflowers.

As Emily ventured deeper into the garden, she noticed something unusual. The plants and creatures within seemed to come to life in her presence. The flowers turned to face her, and the birds sang melodies that filled her heart with joy. It was as though the garden itself recognized her as its new caretaker.

Emily spent her days tending to the garden, caring for the plants and conversing with the animals. She soon discovered that each flower and creature held a story to tell, a secret from the past. The roses whispered of love stories, the oak tree had witnessed ancient battles, and the fish in the pond knew of long-forgotten legends.

One day, as Emily was pruning a rosebush, she uncovered a hidden stone path that led to a magnificent gazebo at the heart of the garden. In the center of the gazebo stood a peculiar, ornate keyhole. With a sense of curiosity, she began her quest to find the key that would unlock its mysteries.

The garden, sensing Emily's desire, guided her on her journey. Each day, she would explore a different part of the garden, and with each new discovery, she would get a little closer to finding the elusive key. She met a wise old tortoise who shared riddles and a mischievous squirrel that led her to hidden clues. Every plant and creature was eager to help her unravel the garden's secrets.

As Emily ventured deeper into the garden's mysteries, she encountered a magical fountain. The fountain's waters shimmered with an otherworldly light, and it spoke to her in a gentle, melodic voice. It told her that the key to the gazebo lay not in a physical form, but in the experiences and stories she collected from the garden.

Emily understood. The key to the gazebo was not a physical object; it was the wisdom, love, and wonder she had gained from her time spent in the garden. Over time, she learned about love, bravery, friendship, and the beauty of the world around her.

Finally, after what felt like a lifetime of exploration, Emily stood before the gazebo once more. As she approached, the keyhole began to glow with a soft, warm light. She touched the keyhole, and it opened, revealing the enchanting wonders of the gazebo.

Inside, she found a magical book filled with the stories of the garden, written in words that seemed to dance on the pages. It was a treasure trove of the enchanting tales and secrets the garden had shared with her. As she read the stories, she felt a deep connection to the garden and the living things within.

Emily realized that the true enchantment of the garden was not just in its physical beauty but in the stories and wisdom it held. She continued to care for the garden, sharing its stories with her family and the world. The garden, now even more vibrant and alive, flourished under her care.

Years passed, and Emily grew into a wise, kind-hearted woman. The enchanted garden became a place of wonder, where people from all walks of life could visit to find solace, learn the wisdom of the natural world, and hear the stories of the flowers and creatures within.

And in the heart of the garden, the gazebo stood as a symbol of the magic that could be found when one listens to the world around them and cherishes the stories it has to share.