The **Media Readability Testbed (MRT)** served as the foundation of this study, providing a custom-built, open-source platform specifically designed to examine digital readability among older adults. The MRT was developed to enable systematic manipulation of key variables, including font type and the presence or absence of hyperlinks, while ensuring consistency in how stimuli were presented across all participants. This platform ensured that experimental conditions remained standardized, reducing variability and enhancing the reliability of results.

In addition to controlling for visual presentation, the MRT incorporated tools for measuring reading-related outcomes such as speed, comprehension, engagement, fatigue, and reading pleasure. The platform also integrated mechanisms for data collection, allowing for the seamless recording of participants' responses and behavioral metrics. By centralizing these capabilities, the MRT streamlined the experimental process, ensuring that all components from visual stimuli to participant feedback were cohesive and systematically implemented. This approach not only enhanced the study's methodological rigor but also provided a robust framework for analyzing how digital readability impacts the experiences of older adults.

**Memory Test**

The memory test was designed to evaluate participants' ability to retain information using word lists of varying lengths (6, 12, and 18 words). Participants were given 30 seconds to memorize each list, followed by five multiple-choice questions to assess their recall. To ensure consistency and fairness, specific types of words were excluded, such as those ending in "s" (unless preceded by "I" or "u"), words ending in "ae," and verbs ending in "ed" or "ing." These exclusions ensured that the word lists were linguistically balanced and minimized potential biases, offering a reliable method for assessing memory performance in older adults.

**Visual Acuity Test**

The visual acuity test assessed participants' ability to recognize letters or symbols at various distances using a standardized chart. Participants were instructed to sit 14 inches away from the screen and use their daily corrected vision method, such as glasses or contact lenses, to complete the task. This setup ensured a standardized testing environment that accounted for individual variations in vision correction. By measuring visual acuity under consistent conditions, this test provided a baseline understanding of participants’ visual capabilities, which was critical for interpreting their performance in subsequent tasks.

**Contrast Sensitivity Test**

Contrast sensitivity, which evaluates the ability to discern differences in shades or contrast, was measured using a standardized contrast chart. Similar to the visual acuity test, participants were instructed to sit 14 inches from the screen and use their corrected vision method. This methodology offered precise insights into participants’ visual perception, particularly their ability to process visual information with varying levels of contrast. By maintaining consistency in the testing protocol, the contrast sensitivity test ensured reliable results that complemented the visual acuity measurements.

**Engagement**

Engagement was measured through a combination of custom-designed questions and a standardized framework. The custom questions focused on participants' immediate engagement with each newspaper article, assessing aspects such as enjoyment, familiarity with the topic, and learning new information. The standardized questions explored broader dimensions of reading engagement, including motivation, self-regulation, and personal beliefs about reading. To ensure consistency in scoring, two negatively phrased items were reverse coded so that higher scores always indicated greater engagement. This dual approach provided a comprehensive view of how participants interacted with and responded to the reading materials, capturing both task-specific and general engagement metrics.

**Fatigue**

Fatigue was measured using a streamlined visual fatigue questionnaire that focused on four key items: "My eyes feel tired," "I feel tired," "At the present, how much eyestrain do you feel?" and "I find it difficult to focus my vision." Items related to physical symptoms such as dry eyes, nausea, and headaches were excluded to reduce participant burden and ensure relevance to the reading task. The remaining items emphasized general tiredness and visual strain, providing a targeted assessment of participants’ subjective fatigue during the reading process. This refined approach maintained the validity of the fatigue measurement while aligning it with the study's focus on readability.

**Reading Pleasure**

Reading pleasure was assessed by evaluating participants’ preferences for fonts and hyperlinks using Likert scales. Font preferences were rated on a scale from 1 to 6, with options ranging from "I strongly prefer Georgia" to "I prefer Roboto," and an additional option for "I don’t prefer any of the font options." Similarly, hyperlink preferences were rated on a scale from 1 to 4, with responses such as "I strongly prefer hyperlinks" and "I do not prefer hyperlinks," along with an option for "I strongly do not prefer hyperlinks." These numerical ratings were directly entered into the statistical analysis to capture the nuances of participants’ preferences. By aligning these preferences with the reading conditions they experienced, this metric provided insights into how subjective enjoyment and preferences influenced the reading experience.

**Reading Speed**

Reading speed was calculated in words per minute (WPM) using the formula WPM=w×60/s, where ww represents the total number of words in the article and ss represents the time spent reading in seconds. This metric provided a standardized measure of how quickly participants read the articles under various conditions, offering insights into how factors like font type and hyperlinks impacted their reading efficiency.

**Comprehension**

Comprehension was measured as the percentage of correct answers participants provided to five multiple-choice questions for each newspaper article. Each participant read six articles, with two articles presented in each font condition (one with hyperlinks and one without). The comprehension score for each article was calculated by dividing the number of correct answers by the total number of questions, expressed as a percentage. This metric allowed for a detailed analysis of how font type and the presence or absence of hyperlinks influenced participants' understanding of the reading material.

**Validity Check**

To ensure data quality, two validity questions were included at the end of the experiment. Participants were asked about their attentiveness and whether they had provided thoughtful responses. Responses indicating inattentiveness or randomness (e.g., “Neutral” or lower for attentiveness, or admitting to providing random answers) resulted in participant exclusion. This process safeguarded the reliability of the study’s findings by ensuring that only data from engaged and attentive participants were included in the analysis.