Title

The Impact of Technology and Presentation Mode on Reading Comprehension Among Blind and Sighted Individuals

Permalink

https://escholarship.org/uc/item/7239h093

Author

Stepien-Bernabe, Natalie Nicole

Publication Date

2019

Peer reviewed|Thesis/dissertation

The Impact of Technology and Presentation Mode on Reading Comprehension Among Blind and Sighted Individuals

by

Natalie Nicole Stepien-Bernabe

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Vision Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Deborah Orel-Bixler, Chair
Professor Dennis Levi
Professor Michael Ranney

Spring 2019

The Impact of Technology and Presentation Mode on Reading Comprehension Among Blind and Sighted Individuals

Abstract

The Impact of Technology and Presentation Mode on Reading Comprehension Among Blind and Sighted Individuals

by

Natalie Nicole Stepien-Bernabe

Doctor of Philosophy in Vision Science

University of California, Berkeley

Professor Deborah Orel-Bixler, Chair

Rapid advances in technology are facilitating the electronic distribution of information, especially via auditory formats. These methods of acquiring information are prevalent in educational settings, such as the use of audiobooks. While auditory formats may be more convenient and economical, they may not be the most beneficial for comprehension ability. In fact, previous research has shown that reading rather than listening leads to superior comprehension among sighted individuals. Two possible explanations for this benefit in comprehension exist: (1) Visual processing is fundamental for reading comprehension, or (2) Reading text is more physically-engaging and effortful than listening, leading to better comprehension ability. This dissertation presents research that contributes to this debate by comparing reading and listening comprehension between sighted individuals and blind, fluent braille readers.

The findings from this research clarify how distributing reading material using different technologies and presentation modes affects comprehension ability. These results are particularly essential for the blind community because previous research on this topic is severely lacking, and currently-used assistive technology primarily presents reading materials in auditory formats. Furthermore, this research is the first to create an assessment that accurately measures comprehension ability among two distinct participant populations, allowing for comparisons between blind and sighted individuals. The benefit of making such comparisons for elucidating the neural underpinnings of cognitive processes, such as comprehension, will be discussed as a novel methodological technique.

I wish to dedicate this dissertation to the late Dr. Valerie Morash.

Dr. Morash introduced me to the field of visual impairment and took me under her wing when I decided to change research directions. She was an amazing mentor and friend. I feel privileged to be able to continue her legacy through conducting research as is reported in this dissertation. The original concept addressed by the studies presented in this report is credited to her.

Contents

List of Figures

List of Tables

Acknowledgments

I am wholeheartedly grateful for the extensive support I received from my advisor, Dr. Deborah Orel-Bixler. I am thankful for her encouragement during my career as a graduate student and for being my musical, symphony, travel, bubble tea, and lunch companion; I will forever cherish those memories!

I express my heartfelt gratitude for the boundless love and support provided by my husband, Lester Bernabe, who had to endure my numerous scientific discussions and talks, emotional struggles, and stress throughout this process.

Words cannot express how grateful I am for my parents, without whom pursuing a higher education and my aspirations would not be possible.

I also thank my sister, Hollie, and all of my friends for encouraging my passions and keeping my head up. A special thanks to Stephanie Wan for the numerous dinner and lunch dates and for her consistent support. Additionally, I am grateful for my best friend, Rachel Meyer, for her encouragement and for taking time to visit me frequently throughout my time in graduate school. Our adventures together will always remain in my best of memories.

Many thanks to my dissertation committee, Dr. Dennis Levi and Dr. Michael Ranney, for providing helpful feedback and insightful comments on my dissertation.

Finally, I thank my many collaborators: Dr. Amanda McKerracher, Daisy Lei, and Dr. Sile O'Modhrain. I am very grateful for their many intellectual contributions. A special thank you to everyone at the Smith-Kettlewell Research Institute - especially Drs. Marilyn Schneck, Gunilla Haegerstrom-Portnoy, James Coughlan, Lori Lott, Suzanne McKee, John Brabyn, and Preeti Verghese - all of whom not only contributed to my academic growth, but also financial and emotional support during a very difficult time. Everyone at the Institute was like a family to me.

Chapter 1

General Introduction

Blindness and visual impairment affect millions of individuals worldwide. In the United States, most individuals who suffer from visual impairments are above 65 years of age. Their visual impairments are generally age-related conditions that cannot be fully corrected with eyeglasses or surgical intervention. In addition to affecting older adults, visual impairments also affect children – particularly, those who are born prematurely (Allen, 2008). Visually-impaired children experience difficulties reading and must use accessible materials to help them navigate the environment, engage in social activity, and persevere in school. Due to advancements in medical care and technology, premature babies have higher survival rates, leading to a large increase in the number of children with visual impairments. Consequently, more children with blindness or visual impairment are attending schools, thereby emphasizing the need to better understand how to distribute educational information in a way that enhances comprehension. This topic will be explored in this dissertation.

1.1 Brief overview of visual impairments and blindness

1.1.1 Definition of legal blindness

According to the National Federation for the Blind (2012), the definition of “legally blind” is the clinical diagnosis of central field acuity of 20/200 or less in the better of the two eyes with the best possible spectacle correction. This acuity signifies that if an object is 200 feet away, one would have to stand 20 feet away from it to see it clearly; whereas a person with normal vision would be able to stand 200 feet away to see it perfectly. Another definition of “legal blindness” per the National Federation for the Blind is the clinical diagnosis of a visual field of 20 degrees or less. All blind participants tested in the experiment presented in Chapters 5 and 6 of this dissertation qualified as legally blind.

1.1.2 Prevalence of visual impairments and blindness

According to the World Health Organization (2016), as of 2014, approximately 285 million people are visually impaired worldwide, 39 million of whom are blind and 246 million of whom have low vision. The most common cause of moderate and severe visual impairment is uncorrected refractive errors (i.e., myopia, hyperopia, and astigmatism) (Dandona & Dandona, 2001; World Health Organization, 2016). Indeed, the prevalence of myopia is increasing over time; however, the reason for this is still unknown. Cataracts are the second most common cause of blindness in middle- and low-income countries, followed by glaucoma (Kingman, 2004; World Health Organization, 2016). Approximately 37% of children with significant visual impairments can be attributed to preterm births (Allen, 2008). Such impairment occurs because the infant's visual system did not develop fully enough to process visual information after birth (Birch & O'Connor, 2001). Children born preterm or with low birth weight have a higher risk for visual impairment, including cerebral visual impairment (CVI) (Vohr et al., 2000). CVI is the leading cause of bilateral visual impairment in children in industrialized nations and results from an insult to the developing brain (Good et al., 2001). According to the American Foundation for the Blind, 30-40% of cases of visual impairment in children is caused by CVI (Roman et al., 2008).

1.2 Current access to reading materials among blind individuals

The first method developed to make reading materials more accessible for individuals with blindness or visual impairment (BVI) is embossed braille. Braille is an alphabetic, tactile language of raised dots, which is read using the fingertips. Today, in addition to embossed braille, there exist many different assistive technologies that students may use for reading materials in school. Such technologies either present the reading material in auditory or braille formats. This section briefly reviews the history of braille and discusses different assistive technologies used to make reading material accessible to BVI individuals.

1.2.1 Embossed, hard-copy braille

Embossed braille may be created with a slate and stylus method, which is designed to manually punch, or emboss, raised tactile dots into braille format on a page (National Federation for the Blind, 2007). The slate serves as a stencil with which to guide the embossing. While this method is extremely time-consuming and not efficient, it is useful for teaching how to write braille. A more commonly-used method for embossing braille is using a braille printer. When using a braille printer, a document is prepared using a word processing program and a braille font setting on a computer. Then, the computer is connected to the braille printer, and the printer embosses the information on paper using solenoids that control different embossing pins (American Foundation for the Blind, nd). The dimensions for braille paper are 11.5 by 11 inches and 8.5 by 11 inches. According to the National Library Service for the Blind and

Physically Handicapped, Library of Congress (2008), the standard height of braille dots (in the "z" dimension) is 0.483 millimeters, the base diameter of the dots is 1.45 millimeters, the distance from center to center of adjacent dots (horizontally or vertically) in the same cell is 2.337 millimeters, and the distance from center to center of corresponding dots in adjacent cells is 6.223 millimeters. Although extensive research has not been done to validate the optimality of these dimensions, they have been functional for reading throughout history.

There are several advantages and disadvantages of using hard-copy braille to access reading information. Learning to read embossed braille assists with spelling and writing in braille. The spatial layout on the braille page also allows the reader to easily make regressions (i.e., backward movements) to clarify previous information, if desired. However, hard-copy braille is time-consuming and costly to produce, and the braille paper is bulky and rather inconvenient to transport. For example, the fifth Harry Potter novel, Harry Potter and the Order of the Phoenix, amounted to a 13-volume stack per book that was more than one foot high (Armstrong, 2003).

1.2.1.1 History of braille

Although Louis Braille has frequently been credited with the creation of embossed braille, the idea was developed by Charles Barbier, a captain of the French army during the 19th century (D'Andrea, 2009). Barbier conceived of the idea in response to Napoleon Bonaparte's demand for a way to send messages at night without being discovered by others during the French Revolution (Lowenfeld et al., 1968). Barbier created a phonetic system called "Ecriture Nocturne," which utilized 12 tactile dots (two rows of six dots) for each character (Lowenfeld et al., 1968). While this system was introduced to and used by the Institute for the Blind in Paris in 1820s, its complicated symbology and large size emphasized the need for revisions that would make the code more useful (Lowenfeld et al., 1968).

Near Paris around the mid-1820s, a young, blind boy named Louis Braille became familiar with "Ecriture Nocturne" and aspired to create a more accessible tactile reading system. When Braille became a teacher before he turned 20 years old, he tested different tactile dot configurations on his students; this led him to create a tactile alphabetic system with six-dot cells (three rows of two tactile dots), which he continuously revised until 1837 (D'Andrea, 2009; Lowenfeld et al., 1968). This braille code, however, was not recognized and used throughout most of Europe until 1878 (D'Andrea, 2009).

While Europe had generally adopted Braille's system, different tactile codes were being used throughout the United States of America. In 1868, the New York Institute for the Blind began using the "New York Point" code, which used less space than Braille's system (D'Andrea, 2009). Although "New York Point" was adopted as the official system for tactile reading in America in 1871, people continued to create other tactile reading systems (e.g., Boston Line Letter, Philadelphia Line, American Braille) and advocated for their general acceptance throughout the United States (D'Andrea, 2009; Dixon, 2000; Lowenfeld et al., 1968). The effort to agree on a single, uniform code led to the "War of the Dots" in America, which was deemed a "conflict that was acrimonious in the extreme" (Irwin, 2009). Eventually, in 1918, "Revised Grade 1-1/2" braille, which was like the braille alphabet being used in Europe, was decided upon as the uniform tactile reading system in the United States (D'Andrea, 2009).

Even after adopting "Revised Grade 1-1/2" braille, there have been continuous efforts to simplify and improve the braille reading system and to bring British and American braille into closer agreement (D'Andrea, 2009). Such revisions especially include incorporating more contractions, where braille symbols and symbol groups are used to represent certain punctuation, letter combinations (e.g., "ea" and "ount"), and whole words (e.g., "were"). In 1992, the Braille Authority of North America (BANA) and the International Council on English Braille (ICEB) started the Unified English Braille (UEB) Code Project, which strived to develop a braille code that would be acceptable to all English-speaking countries (D'Andrea, 2009). Indeed, in 2004, the UEB code was recognized as "an international standard" and is the code that is being taught and used today (D'Andrea, 2009).

1.2.2 Refreshable braille display

In addition to being embossed on braille sheets, braille could also be presented on a refreshable braille display (Figure 1.1). A refreshable braille display is a hardware device that connects to other technological devices (e.g., computers, smart phones, memory cards) and utilizes electronic pins that translate text into braille (Hackett, 2016). Refreshable braille displays are smaller and more transportable compared to hard-copy braille, and they make electronic information quickly and easily accessible in braille format. Some disadvantages of using refreshable braille displays include their high price (range: \$2,000 to \$50,000), and a trade-off of their convenient size is that they only display one line to a half-page of braille at a time (Russomanno et al., 2015).

Figure 1.1: Example of a refreshable braille display. Image captured by Natalie N. Stepien-Bernabe at the Smith-Kettlewell Eye Research Institute, San Francisco, CA on August 8, 2017.

1.2.3 Talking books

Rather than being displayed in braille format, reading materials could also be made accessible in audio formats. The National Library Service at the Library of Congress provides a free "talking book" service, which consist of human voice-recorded books and magazines. They also provide a digital talking book player free of charge (National Library Service for the Blind and Physically Handicapped at the Library of Congress, nd). Individuals also have the option to download audiobooks to a computer or smart phone and listen to them directly from these devices. Another popular device that is used by BVI individuals to access audiobooks is the DAISY (Digital Accessible Information System) "talking book". This device allows for adding bookmarks, locating pages and sections, and skimming through the information (Dietrich, nd).

1.2.4 Screen readers

A screen reader is an additional assistive technology that presents reading material in the auditory domain. Screen readers are text-to-speech software used on electronic devices that employ a synthesized voice to read aloud text information. The two most commonly-used screen reading software programs are JAWS (Job Access with Speech) and NVDA (Non Visual Desktop Access). The voice speech rate and accent may be adjusted in the software settings. Although the synthesized speech used by JAWS tends to be more articulate than what is used by NVDA, it is rather expensive. NVDA, on the other hand, is free; but, oftentimes, users download extensions to improve the speech quality. Screen readers make accessing material (e.g., books, emails, word documents, etc.) in auditory format extremely convenient. Nevertheless, the pronunciation of especially uncommon, complex words is frequently incorrect.

1.3 Reading comprehension and presentation mode

The advancement in technology and development of the assistive devices described above beg the question of whether their means of accessing reading materials (i.e., auditory or braille format) are optimal for enhancing reading comprehension, which is defined as the process of making meaning from text (or auditory) information (Woolley, 2011). Especially when compared to auditory methods of acquiring information, braille is often considered too slow or too challenging to warrant the distribution of sufficient resources to its teaching (Millar, 1997). Furthermore, many blind computer users prefer to access electronic texts in auditory (speech) format, as it is less expensive compared to braille and requires no additional hardware (Russomanno et al., 2015).

Nevertheless, if accessing reading material in braille generates better reading comprehension compared to listening, then this discovery may encourage the allocation of resources to continue teaching braille. It is known that over 60% of BVI individuals are unemployed; however, those who know how to read braille and use it on a weekly basis are employed at a significantly higher rate than non-braille readers (Bell & Mino, 2015). In fact, there is reason to believe that reading braille may promote superior comprehension compared to listening to audio, as research has shown that sighted individuals experience better comprehension and less frequent mind wandering with more physically-engaging modes of reading (i.e., reading text aloud and silently) compared to passively listening to the text (Varao Sousa et al., 2013).

This dissertation will explore the impact of presentation mode (text or braille vs. audio) on reading comprehension among sighted and blind individuals by addressing three separate aims: (1) To develop a novel assessment tool that could be used to evaluate comprehension amongst both sighted and blind populations; (2) To further explore the hypothesis that tasks requiring greater physical engagement improve comprehension ability; (3) To incorporate modern and popular assistive technologies to access reading materials and assess their impact on comprehension among blind braille readers. The results from this dissertation will inform the distribution of information to enhance reading comprehension and will assist in clarifying the cognitive mechanism underlying superior comprehension ability.

Chapter 2

Comprehension assessment tool development

2.1 Introduction

As mentioned in Chapter 1, one aim of the experiments presented in this dissertation is to further explore the hypothesis that more physically-engaging and cognitively-effortful reading tasks enhance comprehension ability. This hypothesis has potentially been supported by previous research that revealed better comprehension among sighted individuals when they actively read text than when they passively listened to the material (e.g., Varao Sousa et al., 2013). A possibility is, however, that the comprehension advantage for text compared to audio is a domain-specific property of the visual system, rather than a general, cognitive property due to increased physical engagement and effort. One way to test this is by developing an assessment tool that could examine comprehension among both sighted and blind individuals after having them access reading material via different presentation modes. This chapter describes the development of a novel assessment tool that successfully probes blind and sighted reading comprehension ability.

This novel tool aims to improve upon tests that have been used to assess comprehension in previous research. To elaborate, many previously-used comprehension assessments required participants to physically write down everything they could recall either immediately or after a delay (e.g., Green, 1981; Dixon et al., 1982; Kintsch et al., 1977) and to answer multiple choice questions (e.g., Varao Sousa et al., 2013; Rogowsky et al., 2016; Lowenfeld, 1945). Using multiple-choice questions to assess comprehension or recall is problematic, insofar as they probe recognition of correct answers rather than recall, potentially simplifying the comprehension exercise (Birenbaum & Tatsuoka, 1987). The novel comprehension tool discussed here, therefore, utilizes free-response questions, which require individuals to recall specific information from the reading material. It may be the first study to utilize the same comprehension test to make comparisons between blind and sighted participant groups. Previous studies that have explored the effect of presentation modality on comprehension did not adequately ensure that the reading material was appropriate for the participants' reading ability level. For the comprehension assessment created and described here, however, the Flesch Kincaid Readability Test (Kincaid et al., 1975) was performed on all the content to ensure that it was appropriate for all participants.

This chapter describes the development and design of a novel, free-response comprehension assessment that was used for experiments in this dissertation, and it describes several Item Response Theory (IRT), reliability, and validity measures that assess the accuracy of the tool.

2.2 Design

2.2.1 Construct map

The construct measured in this dissertation is comprehension. According to Woolley (2011), comprehension is the process of making meaning from text (or audio) information and gaining an overall understanding of what is described, not just the meaning of isolated words/sentences (Woolley, 2011). A description of the construct map, a map that defines a uni-dimensional idea (in this case, comprehension) and orders both persons and responses from less (near the bottom) to more (near the top) difficult, is provided in Figure 2.1 (Wilson, 2005, p. 6).

Direction of increasing comprehension

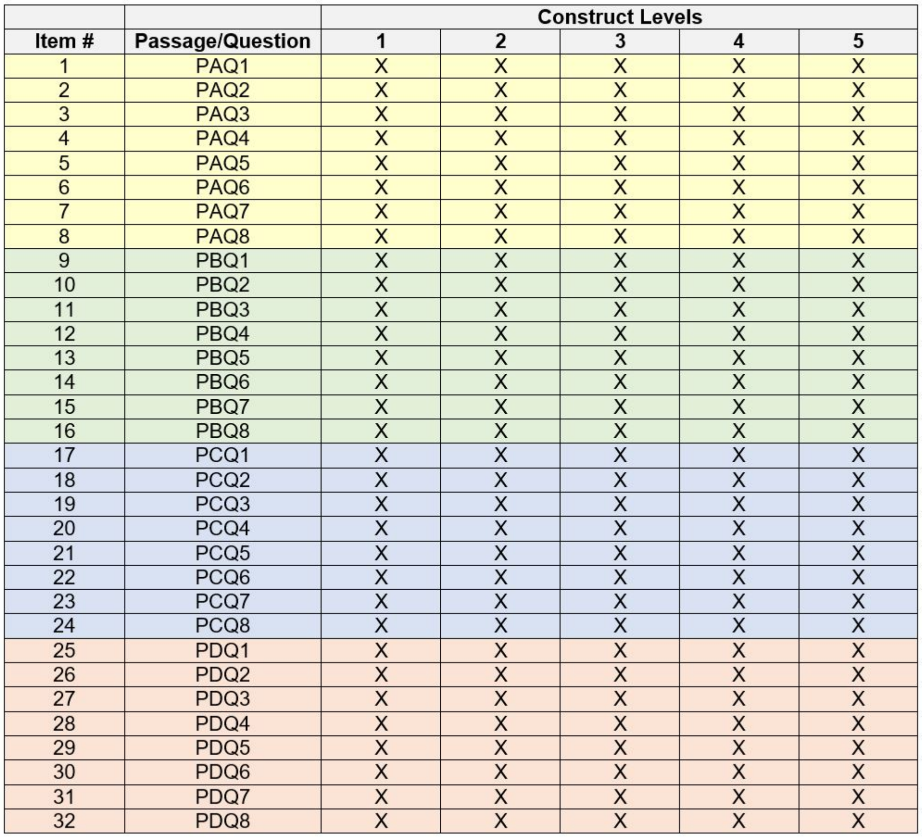
Construct Levels	Response Description	Participant Descriptions
Level 5: Completely Recalling or Inferencing	The response includes all relevant scientific terms or complete application/inference.	Participants remember all terminology from the passage and are able the completely apply and synthesize the information they learned to address a novel scenario.
Level 4: Partially Recalling or Inferencing	The response includes some (but not all) relevant scientific terms or partial (but not complete) application/inference.	Participants can remember some of the terminology used in the passage and are able to synthesize the knowledge they acquired from the passage and partially apply it to address a novel scenario.
Level 3: Describing	Response includes descriptions of relevant scientific terms, but do not explicitly state the required vocabulary to answer the question.	Participants do not remember explicit terminology used in the passage, but are able to describe the terms.
Level 2: Falsely Recalling or Inferencing	Response may state relevant terminology from the passage, but the descriptions/inferences provided are incorrect.	Participants do not recall terms stated in the passage or are unable to make inferences.
Level 1: Missing or Irrelevant Response	Response does not provide any information ("I don't know", "I don't remember"), or provides material that was not related to the passage material.	Participants do not remember information provided in the passage.

Direction of decreasing comprehension

Figure 2.1: Comprehension construct map.

2.2.2 Item development

The initial comprehension assessment tool consisted of four different scientific passages from Prentice Hall's High School Biology eTextbook by Pearson publishing company (Miller & Levine, 2008). Scientific passages were selected as part of the experimental motivation, as blind individuals are particularly underrepresented in STEM-related fields; thus, the findings from the experiments that utilize this tool to assess blind individuals' comprehension will have meaningful and important implications. The Flesch Kincaid Readability Test was performed on each of the passages, which revealed that all passages were written at a 12th-grade reading ability level (Kincaid et al., 1975). Eight polytomous, free-response comprehension questions were developed for each of the four passages (total of 32 items; see Appendix for full list of passages and questions). Half (four out of eight per passage) were literal (i.e., the question could be answered from information stated explicitly in the passage), while the other half were inferential (i.e., the answer is not stated explicitly in the passage, but participants must apply the knowledge acquired from the passage). A graphical representation of the item design is provided in Figure 2.2, where an "X" is placed for the construct levels that each item may cover. Responses from all 32 items could span all five construct levels described in Figure 2.1. The different colors represent the passage for which the item is assessing comprehension.



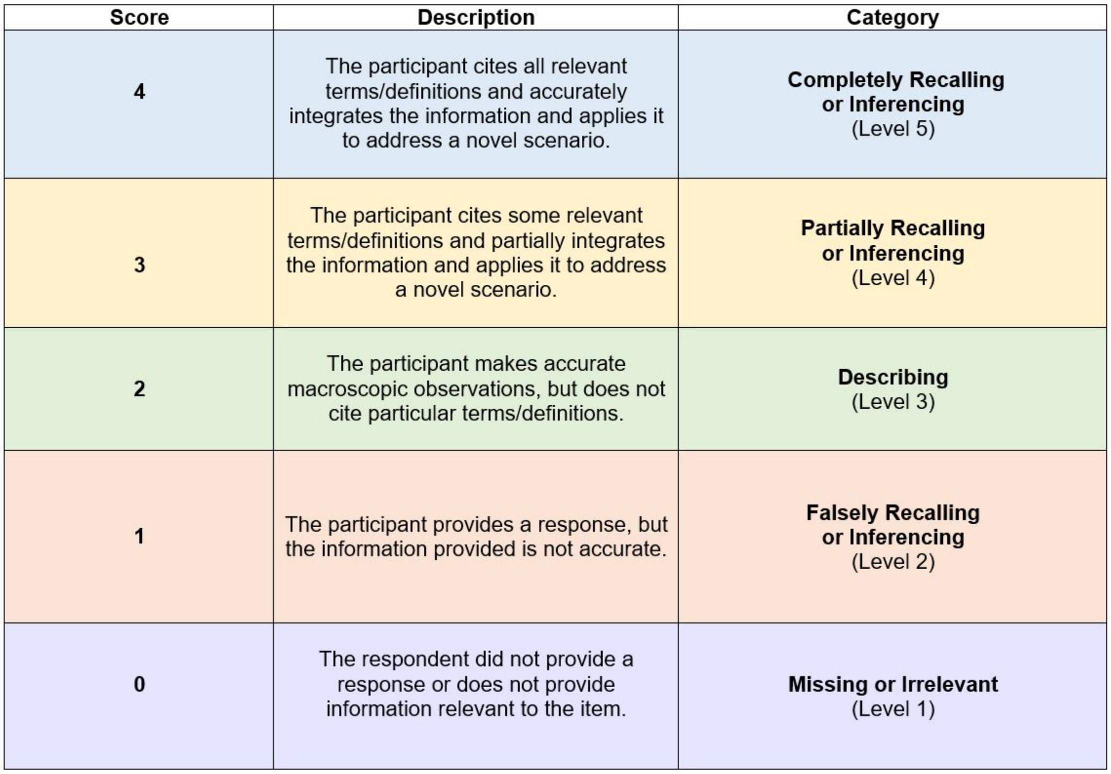
		Construct Levels				
Item #	**Passage/Question**	**1**	**2**	**3**	**4**	**5**
1	PAQ1	X	X	X	X	X
2	PAQ2	X	X	X	X	X
3	PAQ3	X	X	X	X	X
4	PAQ4	X	X	X	X	X
5	PAQ5	X	X	X	X	X
6	PAQ6	X	X	X	X	X
7	PAQ7	X	X	X	X	X
8	PAQ8	X	X	X	X	X
9	PBQ1	X	X	X	X	X
10	PBQ2	X	X	X	X	X
11	PBQ3	X	X	X	X	X
12	PBQ4	X	X	X	X	X
13	PBQ5	X	X	X	X	X
14	PBQ6	X	X	X	X	X
15	PBQ7	X	X	X	X	X
16	PBQ8	X	X	X	X	X
17	PCQ1	X	X	X	X	X
18	PCQ2	X	X	X	X	X
19	PCQ3	X	X	X	X	X
20	PCQ4	X	X	X	X	X
21	PCQ5	X	X	X	X	X
22	PCQ6	X	X	X	X	X
23	PCQ7	X	X	X	X	X
24	PCQ8	X	X	X	X	X
25	PDQ1	X	X	X	X	X
26	PDQ2	X	X	X	X	X
27	PDQ3	X	X	X	X	X
28	PDQ4	X	X	X	X	X
29	PDQ5	X	X	X	X	X
30	PDQ6	X	X	X	X	X
31	PDQ7	X	X	X	X	X
32	PDQ8	X	X	X	X	X

Figure 2.2: Graphical representation of the construct level span for each polytomous free-response item.

2.2.3 Outcome space and scoring guide

An outcome space (i.e., a space that describes how responses are categorized and scored; Wilson, 2005, p.12) was produced for each of the items, and they were combined into one general outcome space (Figure 2.3). In the outcome space below, the scoring guide is related to each level of the construct map.

Like the construct map, the outcome space is divided into distinct categories from the lowest level of comprehension at the bottom to the highest level at the top. Participants are instructed to say, “I don’t know” for any item to which they do not know how to respond and “I can’t remember” for any item to which they cannot remember the relevant information from the corresponding passage; both response types receive a score of 0. Inaccurate responses receive a score of 1. If participants make general, but relevant, observations about the item, but do not explicitly state required terms or definitions from the corresponding passage, then they will receive a score of 2. Participants who succeed in some citing the relevant terms and/or definitions needed to accurately address the item or in partially apply the passage information to make inferences receive a score of 3. A maximum score of 4 is achieved if participants successfully cite all relevant terms and/or definitions, or completely apply passage information to formulate an answer that is not stated verbatim in the passage.



Score	Description	Category
4	The participant cites all relevant terms/definitions and accurately integrates the information and applies it to address a novel scenario.	**Completely Recalling or Inferencing** (Level 5)
3	The participant cites some relevant terms/definitions and partially integrates the information and applies it to address a novel scenario.	**Partially Recalling or Inferencing** (Level 4)
2	The participant makes accurate macroscopic observations, but does not cite particular terms/definitions.	**Describing** (Level 3)
1	The participant provides a response, but the information provided is not accurate.	**Falsely Recalling or Inferencing** (Level 2)
0	The respondent did not provide a response or does not provide information relevant to the item.	**Missing or Irrelevant** (Level 1)

Figure 2.3: Outcome space.

2.3 Data collection

The data collection process is described in detail in Chapter 3; however, a summary is provided here. This comprehension assessment tool was administered to 34 blind individuals (two have been excluded from further analysis due to reasons explained in Chapter 5) and 31 sighted individuals. All participants had at least a high school diploma and were at least 18 years old. A subset of the Wechsler Individual Achievement Test Third Edition was administered to each participant to assess oral reading fluency (Pearson, 2009). During this test, participants were asked to read a series of words aloud. The number of words pronounced correctly represented oral reading fluency ability. It is anticipated that reading fluency score will be positively correlated with comprehension ability, which will be further discussed in Section 2.6.

2.4 Results

This section explains the assessment calibration and steps taken to determine which questions accurately measured comprehension. Importantly, blind and sighted participant responses were analyzed separately, as they constitute separate groups that have cognitive differences that could potentially confound the results if combined.

2.4.1 Item fit for sighted and blind participants

Scores were analyzed utilizing Rasch model estimation with the “CRASCH” package in R 3.4.4 (Arneson et al., 2015). The scores predicted by the Rasch model were compared with the actual scores from the blind and sighted participants, separately, yielding an infit mean square (MNSQ) statistic value for each item. The infit MNSQ value for each item reflects the amount of randomness of the measurement system. Values that are less than 1.0 are considered too predictable, whereas values that are greater than 1.0 are considered noisy. Previous research has indicated that an infit MNSQ between 0.75 and 1.33 is acceptable (Adams & Khoo, 1993; Wilson, 2005) and is represented by the blue shaded regions in Figure 2.4. The infit MNSQ values were calculated for each of the 32 total items for blind and sighted participants, separately. The infit MNSQ values for all 32 items fell within the acceptable range for sighted participants (Figure 2.4, left); but, two of the items fell outside of the range for blind participants (PAQ3=0.73 and PDQ7=1.34; Figure 2.4, right). Consequently, these items were removed from further analysis for both blind and sighted individuals to equate the items across both groups. Additionally, because the two misfit items were of different passage topics and question types (literal and inferential), two items (one literal, one inferential) from each passage were removed from each passage. To determine which items to remove, the difference between each remaining item and the value of 1.0 was calculated. This procedure was done with the data from blind and sighted participants, and all items with the largest difference came from the blind participant data. These eight items were removed, leaving a total of six questions (three literal, three inferential) per passage. The infit MNSQ values were computed for the final 24 best-fitting questions and assessed in the sections below to ensure fit for both participant groups.

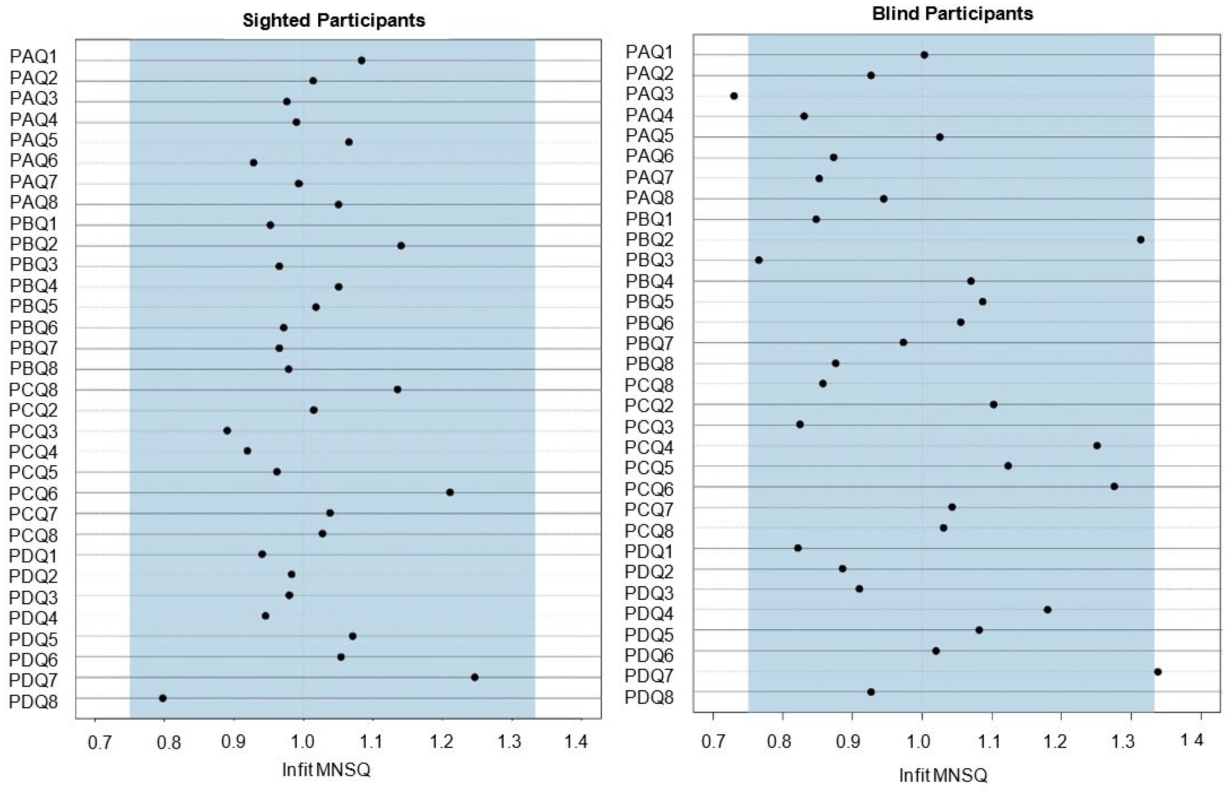


Figure 2.4: Infit mean square values for (left) sighted and (right) blind participants.

2.4.2 Revised item fit for sighted and blind participants

The infit MNSQ values for the fitted 24 items described above are shown for sighted and blind participant data in Figure 2.5. All values existed within the acceptable range (shaded in blue in Figure 2.5) between 0.75 and 1.33. These items, therefore, will be used for all further measurements in this section and for all data analysis throughout this dissertation.

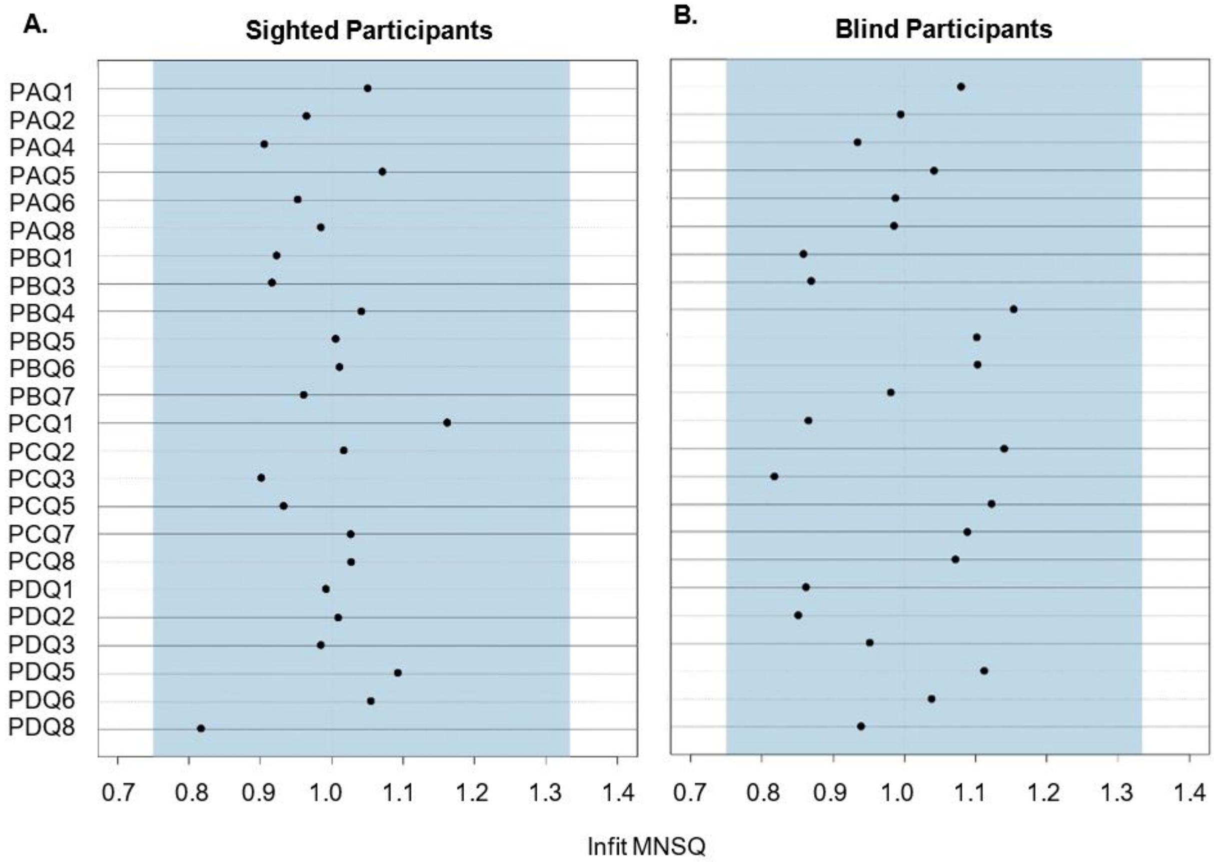


Figure 2.5: Revised infit mean square values for (A) sighted participants and (B) blind participants.

2.4.3 Respondent fit for sighted participants

Respondent fit statistics describe the relationship between the responses observed from a particular participant and the responses that are expected from the Rasch model for that participant (Chien et al., 2011). Smaller infit statistics (MNSQ) signify a closer relationship between the observed and expected responses. Similar to the infit MNSQ values for assessing item fit, when these values are used for evaluating respondent fit, they signify the amount of randomness that occurs in each participant's responses to the assessment tool. In addition to using the infit MNSQ statistic to investigate respondent fit, a visual diagram called a Kidmap may be used. Kidmaps provide information about each participant's observed and expected (i.e., estimated comprehension ability) response patterns for each item, which is represented by a dot plotted on the map (Chien et al., 2011). By comparing the expected and observed responses for each item using a Kidmap, one may determine which items were too easy or too difficult given each participant's estimated comprehension ability level. The following sections provide and explain these methods for assessing respondent fit for the sighted and blind participants, respectively.

A scatter plot was produced to compare sighted participants' estimated ability based on their scores with their infit MNSQ values. According to Figure 2.6, the participant with the highest fit index had an infit value of 1.669 (participant #3), and the participant with the lowest fit index had an infit value of 0.479 (participant #17). Furthermore, the respondents with the highest score had an estimated ability value of 0.571 (participants #2 and #10), and the respondent with the lowest score had an estimated ability value of -0.608 (participant #16). These values are marked with a green circle in Figure 2.6.

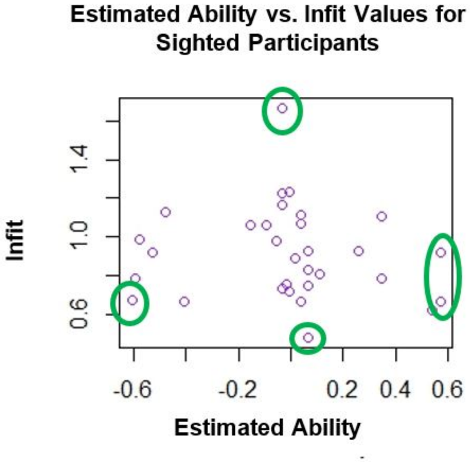


Figure 2.6: Estimated comprehension ability versus infit statistics for sighted participants. The green circles denote maximum and minimum values for each variable.

Separate Kidmaps were generated for the participants with the highest fit index value (participant #3), best score (participants #2 and #10), and worst score (participant #16). If an item falls within the upper-left yellow zone of the Kidmap, this indicates that it is considered too difficult for the participant's comprehension ability level (represented by the horizontal black line), but the participant answered it correctly. An item that falls within the lower-right yellow region should be too easy for the participant's comprehension ability level, but the participant answered it incorrectly. This analysis, therefore, provides additional information regarding how well separate items assess each participant's comprehension.

Figure 2.7 shows a Kidmap for participant #3, who had the highest fit index value, and for participant #17, who had the lowest fit index value. These graphs show that all items fall outside of the yellow regions; this result indicates that these respondents got all the questions correct or incorrect that they should have based on their respective comprehension ability levels.

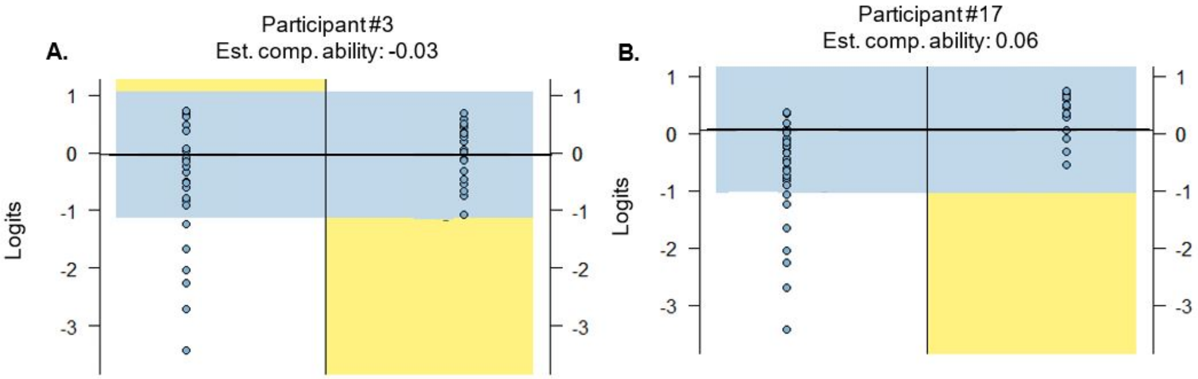


Figure 2.7: Kidmaps for the sighted participant with the (A) highest and (B) lowest fit index value.

A Kidmap for the participants with the best score are shown side-by-side in Figure 2.8. For participants #2 and #10, items PAQ5 and PDQ6 are respectively concerning, because these items are below the participants' ability levels, but they were not answered correctly.

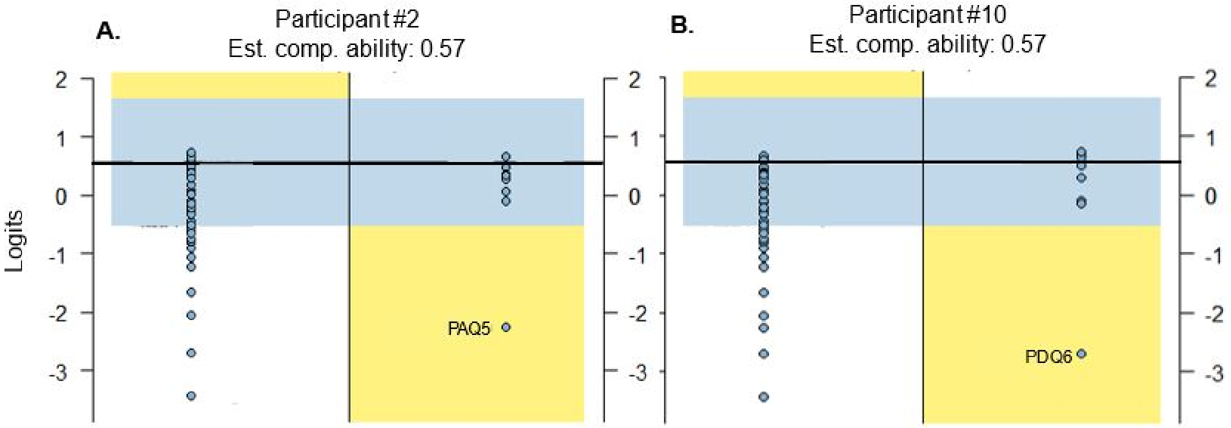


Figure 2.8: Kidmaps for the sighted participants with the highest estimated comprehension ability (A) Participant #2 and (B) Participant #10.

Finally, the Kidmap for participant #16 (worst score) is shown in Figure 2.9. For this participant, item PBQ7 fell outside of the yellow region. This participant answered item PBQ7 incorrectly, even though he or she should have answered it correctly based on his or her comprehension ability level.

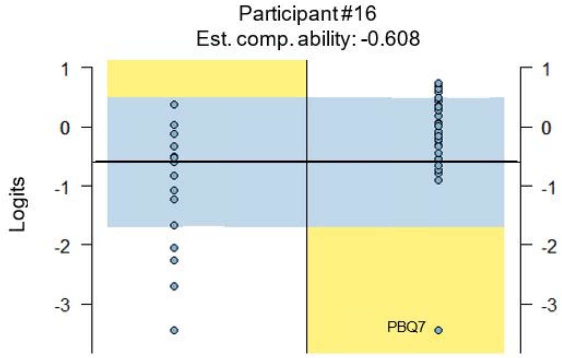


Figure 2.9: Kidmap for the sighted participant with the lowest estimated comprehension ability, Participant #16.

Altogether, these findings reveal that for participants at the extremes (i.e., those with maximum and minimum fit statistics and scores), the majority of the items appropriately assess comprehension ability.

2.4.4 Respondent fit for blind participants

As with the sighted participants, a scatter plot was generated to compare blind participants' estimated comprehension ability based on their scores with their infit MNSQ values. According to Figure 2.10, the participant with the highest fit index had an infit value of 1.607 (participant #29) and the participant with the lowest fit index had an infit value of 0.304 (participant #10). Furthermore, the participant with the highest score had an estimated ability value of 1.534 (participant #3), and the participant with the lowest score had an estimated ability value of -1.212 (participant #10). These values are marked with a green circle in Figure 2.10.

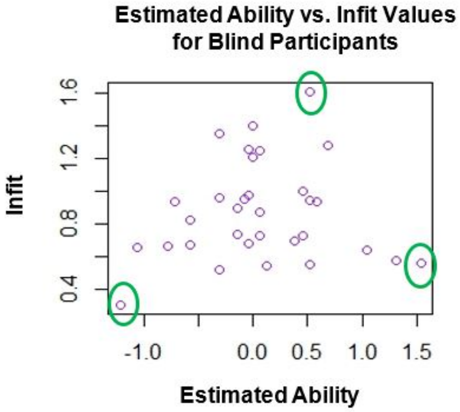


Figure 2.10: Estimated comprehension ability versus infit statistics for blind participants. The green circles denote maximum and minimum values for each variable.

Figure 2.11 shows a Kidmap for participant #31, who had the highest fit index value, and participant #10, who had the lowest fit index value. The graphs reveal that participant #31 should have gotten two items (PDQ6 and PBQ5) correct based on his or her comprehension ability level, but got them incorrect. All other items for this participant and all items for participant #10 behaved as expected given each participant's comprehension ability level.

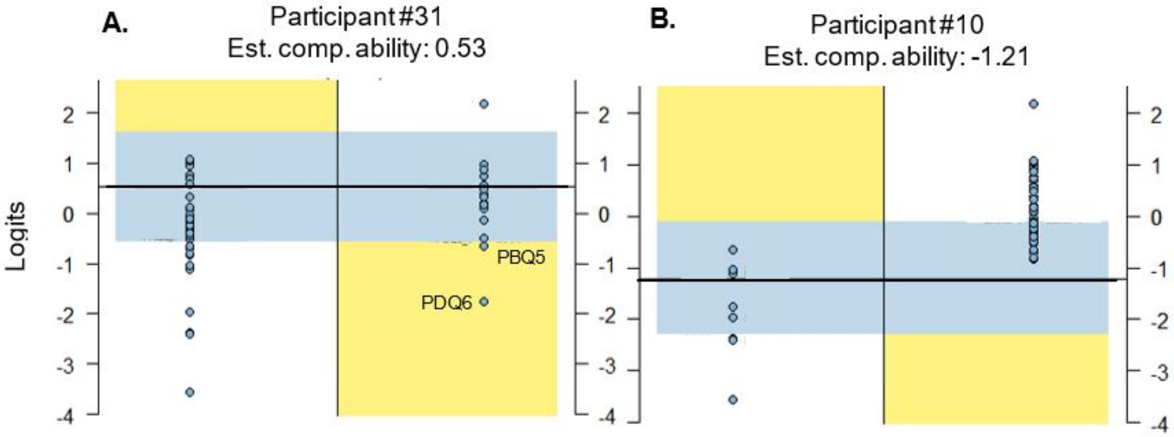


Figure 2.11: Kidmaps for the blind participants with the (A) highest and (B) lowest fit index values. Participant #10 also had the lowest comprehension ability.

Kidmaps were also produced for blind participants with the best score (see Figure 2.12 for participant #3; see Figure 2.11 for participant #10). Whereas there are no concerning items for participant #10, one item for participant #3 fell within the yellow region, indicating that this participant answered this item incorrectly even though he or she should have answered it correctly based on his or her ability level.

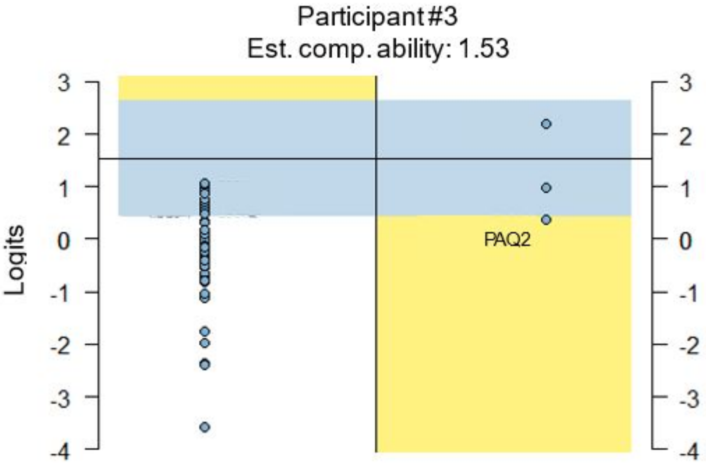


Figure 2.12: Kidmap for the blind participant with the highest comprehension ability.

In summary, as with the sighted participant Kidmaps, most items for these blind participants fall outside of the two yellow regions, meaning they are appropriate assessments of differing comprehension ability levels and fit statistics.

2.4.5 Calibration of the assessment for sighted and blind participants

To calibrate the comprehension assessment, a Wright map was produced for the sighted and blind participant data (Figure 2.13, see following page). A Wright map was generated using the "CRASCH" package in R 3.4.4, which relates the experimental data from blind and sighted participants, separately, to the construct map (Wilson, 2005, p.96).The first threshold should be below, or easier, than achieving the other threshold levels; the second level should be below the third and fourth; and the third should be below the fourth. Although this pattern is not shown explicitly for all items, the mean threshold for each level across items for sighted and blind individuals follow the anticipated increasing trend (Sighted: -1.14, 0.092, 0.333, 0.603; Blind: -0.837, 0.383, 0.239, 1.037). These findings signify that on average, it is more difficult for participants to achieve a higher score than a lower score, which means that the construct map developed in Section 2.2.1 and the items are reasonably acceptable at describing the different comprehension levels assessed by this measurement tool.

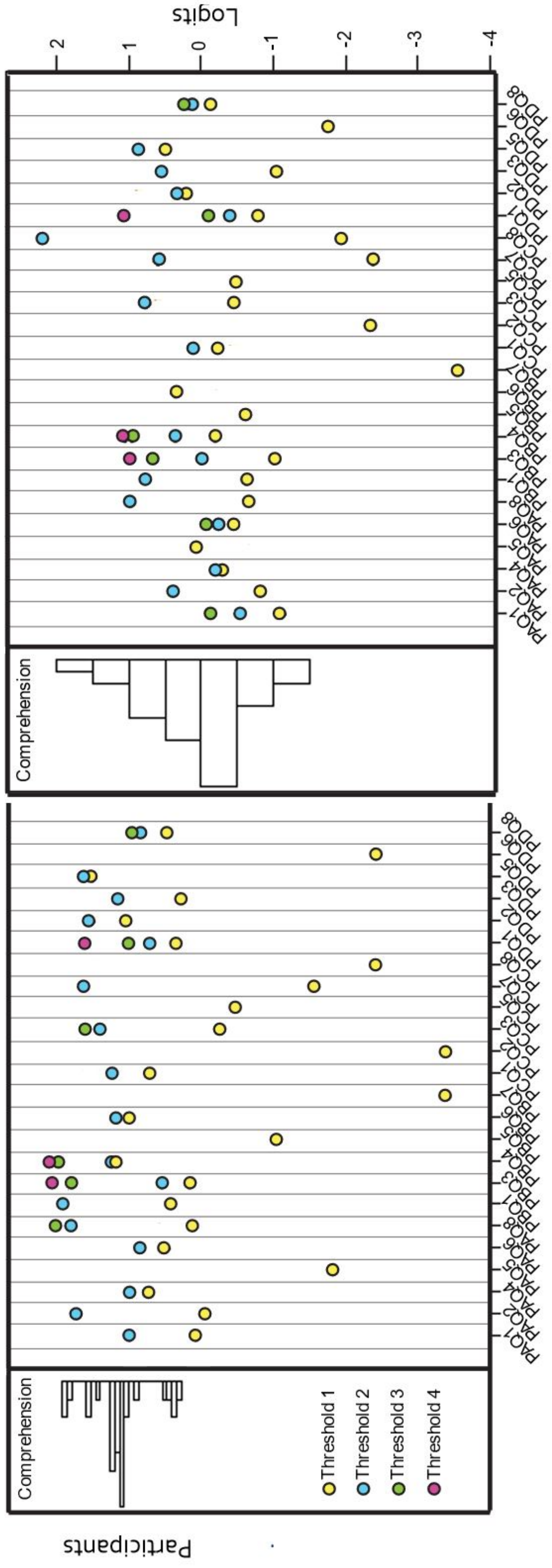


Figure 2.13: Wright maps for (left) sighted participants and (right) blind participants.

2.5 Reliability measures

2.5.1 Internal construct reliability

One graphical representation of how reliably the measurement tool assessed comprehension abilities is the Test Information Curve (TIC; Figure 2.14). The Information value is equal to the reciprocal of the square of the standard error of the mean (Lord, 1980). Consequently, higher values on the TIC correspond to higher information values, whereas lower values correspond to lower information values. The logit values on the x-axis represent the estimated comprehension ability levels of the participants predicated by Rasch modeling. Figure 2.14 reveals that the most sensitive part of this instrument for sighted participants is clustered between approximately -0.6 and 0.57 logits and for blind participants is between about -0.78 and 0.69 logits. As the vast majority of blind and sighted participants fall within these respective ranges, the TIC graphs indicate acceptable reliability.

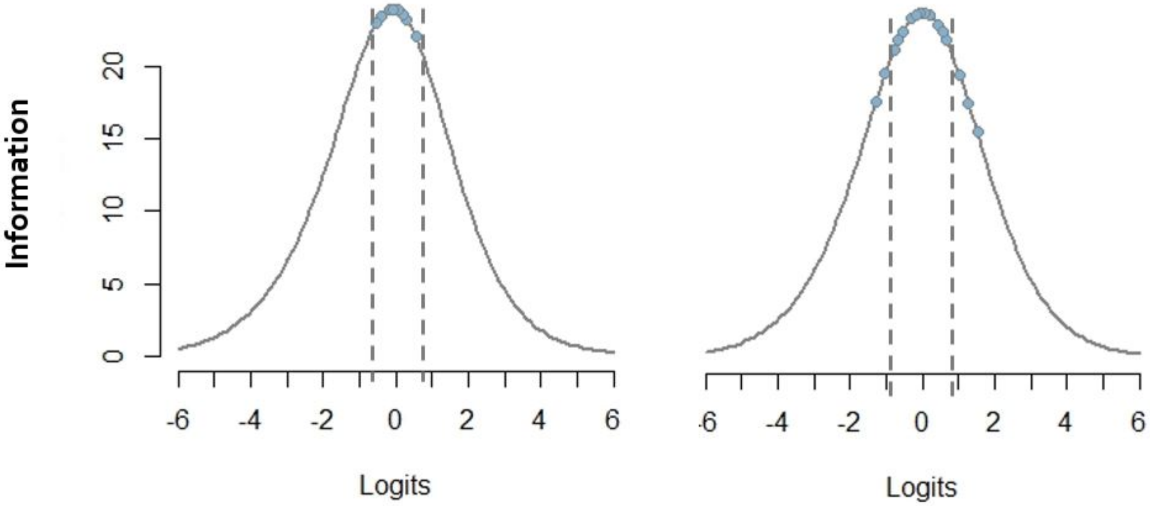


Figure 2.14: Test information curves for (left) sighted participants and (right) blind participants.

Another way to assess reliability of this measurement tool is to compute the coefficient alpha (i.e., Cronbach's Alpha) for polytomous responses (Cronbach, 1951; Wilson, 2005, p. 138) and the Person Separation Reliability value. These values range between 0 and 1, with a value closer to 1 signifying that the items in the assessment tool likely measure the same underlying concept of comprehension. For blind participants, the Cronbach's Alpha value is 0.822 and the Person Separation Reliability value is 0.852. For sighted participants, the Cronbach's Alpha value is 0.658 and the Person Separation Reliability value is 0.735. Altogether, these values indicate the instrument acceptably assesses comprehension (Nunnally, 1978; Hinton et al., 2004; Chang et al., 2014).

2.5.2 Inter-rater reliability

Two individuals (Natalie Stepien-Bernabe and Daisy Lei, a research scientist at the Smith-Kettlewell Eye Research Institute) independently graded all participant responses (including the questions that were deleted from final data analysis) from the first 19 participants for the blind and sighted individuals (32 questions (before deletions) * 19 individuals = 608, for blind/sighted data, separately).

Table 2.1 displays the inter-rater consistency for sighted participants, and Table 2.2 shows the inter-rater consistency for blind participant data. Rater 1 (myself) is in the first column with Rater 2's (Daisy Lei's) ratings at the heads of the next four columns. This display layout was adopted from (Wilson, 2005, p. 142). There are five possible score levels (0, 1, 2, 3, or 4), and the number of occurrences of each possible pair is recorded in the center of the table. Recall that:

- A score of 0 indicates the participant does not provide a response or does not provide relevant information.
- A score of 1 represents that the participant provides a response, but the information is not accurate.
- A score of 2 is given when participants make macroscopic observations, but do not cite relevant terms and/or definitions.
- A score of 3 indicates that the participant cites some of the relevant terms and definitions, and partially integrates them to address a novel scenario.
- A perfect score of 4 represents when participants cite all relevant terms and definitions and completely integrate and apply them to a novel scenario.

The index of agreement is equal to the proportion of exact agreement between the raters (Wilson, 2005, p. 142). This is calculated by taking the sum of the diagonal entries in Tables 2.1 and 2.2 and dividing this sum by the total value in the bottom right corner of the table. Specifically, this proportion was calculated by: *Sighted:* (137+6+108+10+340)/608 = **98.84%**; *Blind:* (184+17+91+12+284)/608=**96.71%**, which is very high agreement of scores between the two raters.

Sighted Participants	**Rater 2**					
Rater 1	**0**	**1**	**2**	**3**	**4**	**Total**
0	137	0	0	0	0	137
1	0	6	0	0	0	6
2	3	0	108	0	2	113
3	0	0	0	10	0	10
4	1	0	1	0	340	342
Total	141	6	109	10	342	608

Table 2.1: Inter-rater consistency for sighted participants.

Blind Participants	**Rater 2**					
Rater 1	**0**	**1**	**2**	**3**	**4**	**Total**
0	184	0	1	0	3	188
1	2	17	1	0	0	20
2	1	0	91	2	4	98
3	0	0	0	12	0	12
4	1	0	3	2	284	290
Total	188	17	96	16	291	608

Table 2.2: Inter-rater consistency for blind participants.

As a separate measure of inter-rater reliability, the Cohen's Kappa value (k) was calculated ("Cohen's Kappa Index of Inter-rater Reliability", n.d.), which is a statistic that measures inter-rater agreement while considering the possibility of the agreement across raters occurring by chance (Fleiss et al., 1969). The possibility was calculated for each score (0, 1, 2, 3, and 4; see Tables 2.3 and 2.4) by calculating the proportion of the total of a particular score from each rater, multiplying these proportions to get a total proportion value, and then finding the product of this total proportion value and the total number of responses. For example, to calculate the amount of responses of score 0 that occurred by chance, the following computational steps were performed: Rater 1: 137/608 = 0.225 Rater 2: 141/608 = 0.232 0.225*0.232=0.0522 of 5.22% of the responses with score of 0 are due to chance 5.22% of 608 total responses = **31.74 responses** These same steps were performed for each score level to achieve the number of response scores that were agreed on due to chance (Tables 2.3 and 2.4).

Sighted	**0**	**1**	**2**	**3**	**4**	**Total**
Agreement	137	6	108	10	340	601
By Chance	31.74	0.06	20.26	0.16	77.93	130.15

Table 2.3: Number of sighted participants' scores agreed upon due to chance.

The k statistic was calculated by using the following formula:

$$\frac{\text{Number of Agreements - Number of Agreements Due to Chance}}{\text{Number of Total Responses - Number of Agreements Due to Chance}} \frac{601-130.15}{608-130.15} = \frac{470.85}{477.85} = \mathbf{0.984}$$

Blind	**0**	**1**	**2**	**3**	**4**	**Total**
Agreement	184	17	91	12	12	588
By Chance	58.13	0.56	15.48	0.32	138.8	213.29

Table 2.4: Number of blind participants' scores agreed upon due to chance.

$$\frac{\text{Number of Agreements - Number of Agreements Due to Chance}}{\text{Number of Total Responses - Number of Agreements Due to Chance}} = \frac{588-213.29}{608-213.29} = \frac{374.71}{394.71} = \mathbf{0.949}$$

Since k=0.984 (sighted) and 0.949 (blind), this means that after subtracting out the agreement due to chance, there is agreement between the two raters approximately 98.4% and 94.9% of the time for blind and sighted scores, respectively; therefore, there is a very good level of agreement between the two raters.

2.6 Validity measures

2.6.1 Evidence based on instrument content

Evidence based on instrument content is provided in Sections 2.2.2-2.2.4 and Section 2.4.5 of this chapter. That is, the relationship between the instrument's content and the construct (i.e., comprehension ability) was documented with the "Four Building Blocks" approach (Wilson, 2005, p. 17). In particular, (1) the construct and construct levels (respondent and response levels) were defined and explained in the construct map in Section 2.2.1; (2) the items and their design were detailed in Section 2.2.2; (3) the outcome space, scoring procedures, and strategies were developed in Section 2.2.3; (4) data was collected from a study to produce a Wright map and infit statistic values in Section 2.4.5.

2.6.2 Evidence based on internal structure, instrument level

If the results produced by the Wright map reflects the construct map well, then the lowest item threshold would represent Step 1 ("Falsely Recalling/Inferencing"), the next set of item thresholds would represent Step 2 ("Describing"), then Step 3 ("Partially Recalling/Inferencing"), and the highest set of item thresholds would represent Step 4 ("Completely Recalling/Inferencing"). The Wright maps in Figure 2.13 (Section 2.4.5) reveal that, although this pattern does not hold true across all items, it does occur for all individual items and for the average step threshold values, as described in the instrument calibrations section. This finding implies an acceptable relationship between the construct map expectations and the empirical data that was collected.

2.6.3 Evidence based on internal structure, item level

To investigate whether the items are consistent with the comprehension instrument as a whole, a table was produced with the mean person location for each response category and item and the "count value", or number of individuals in each score group, which is placed in parentheses (see Tables 2.5 and 2.6). If the mean location of each group increases as response category/score increases, then this supports the expectation that participants with higher comprehension ability tend to score higher on each item (Wilson, 2005, p. 152). Cells with "NA" mark cases where there were no participant responses that fell within the particular category, so all of these cells have a "count value" of zero. The highlighted cells in Tables 2.5 and 2.6 represent cases where there is a drop in mean person location. Overall, out of 75 total cases (cells with NA not included in the total count), 59 (78.7%) cases for sighted participants and 68 (90.7%)

cases for blind participants followed the anticipated increasing trend across construct levels. The highlighted cells in the two tables indicate that it is easier for individuals to achieve a higher score (fall in a higher level on the construct map) than it is for them to receive a lower score (fall in a lower level on the construct map). Furthermore, for sighted individuals, six out of 16 of these highlighted cases and, for blind individuals, five out of seven of these cases have “count values” less than five. It is possible that increasing the sample size may reduce the amount of error in the mean person location, which may yield mean person locations that follow the expected increasing trend with increasing response categories. In the remaining highlighted cases that deviate from the expected increasing trend, the lower-than-expected mean person locations may be indicative of issues with the scoring, item design, and/or the construct map, rather than being attributed to large error.

Items	**1**	**2**	**3**	**4**	**5**
1 (PAQ1)	-0.10504 (4)	NA (0)	-0.10275 (4)	NA (0)	0.014341 (23)
2 (PAQ2)	-0.1611 (6)	NA (0)	-0.09753 (10)	NA (0)	0.096037 (15)
3 (PAQ4)	-0.23261 (9)	NA (0)	-0.02282 (2)	NA (0)	0.081889 (20)
4 (PAQ5)	0.234412 (3)	NA (0)	NA (0)	NA (0)	-0.04302 (28)
5 (PAQ6)	-0.2139 (6)	NA (0)	-0.27075 (2)	NA (0)	0.057546 (23)
6 (PAQ8)	-0.32377 (8)	NA (0)	0.083403 (15)	-0.03085 (2)	0.149922 (6)
7 (PBQ1)	-0.18464 (10)	NA (0)	-0.01707 (12)	NA (0)	0.172207 (9)
8 (PBQ3)	-0.37566 (6)	-0.18173 (3)	0.101804 (14)	-0.09323 (2)	0.176504 (6)
9 (PBQ4)	-0.09863 (17)	-0.03085 (1)	-0.02482 (9)	0.540313 (1)	0.296443 (3)
10 (PBQ5)	-0.1126 (5)	NA (0)	NA (0)	NA (0)	0.002372 (26)
11 (PBQ6)	-0.14436 (14)	NA (0)	0.003138 (2)	NA (0)	0.100897 (15)
12 (PBQ7)	-0.60778 (1)	NA (0)	NA (0)	NA (0)	0.003548 (30)
13 (PCQ1)	-0.05839 (11)	NA (0)	0.015244 (4)	NA (0)	0.005000 (16)
14 (PCQ2)	0.037129 (1)	NA (0)	NA (0)	NA (0)	-0.01795 (30)
15 (PCQ3)	-0.24683 (6)	NA (0)	-0.1273 (11)	0.012399 (2)	0.195376 (12)
16 (PCQ5)	-0.24132 (7)	NA (0)	NA (0)	NA (0)	0.049497 (24)
17 (PCQ7)	0.349508 (1)	NA (0)	-0.20756 (6)	NA (0)	0.016440 (24)
18 (PCQ8)	0.016805 (2)	NA (0)	-0.01845 (29)	NA (0)	NA (0)
19 (PDQ1)	-0.33479 (7)	0.025765 (3)	-0.00042 (4)	-0.00642 (6)	0.164102 (11)
20 (PDQ2)	-0.09008 (16)	NA (0)	-0.14735 (5)	NA (0)	0.167663 (10)
21 (PDQ3)	-0.20512 (6)	NA (0)	-0.11277 (6)	NA (0)	0.074001 (19)
22 (PDQ5)	-0.05057 (23)	NA (0)	0.066281 (1)	NA (0)	0.085071 (7)
23 (PDQ6)	0.283578 (2)	NA (0)	NA (0)	NA (0)	-0.03684 (29)
24 (PDQ8)	-0.31783 (7)	-0.40625 (3)	NA (0)	0.570674 (1)	0.118579 (20)

Table 2.5: Mean sighted person location for the five response categories.

Items	**1**	**2**	**3**	**4**	**5**
1 (PAQ1)	-0.05231 (3)	-0.99273 (3)	-0.41047 (3)	NA (0)	0.278421 (23)
2 (PAQ2)	-0.47525 (7)	NA (0)	0.153707 (9)	NA (0)	0.248784 (16)
3 (PAQ4)	-0.44716 (9)	NA (0)	0.068601 (1)	NA (0)	0.272409 (22)
4 (PAQ5)	-0.10072 (16)	NA (0)	NA (0)	NA (0)	0.228042 (16)
5 (PAQ6)	-0.445 (8)	-0.28964 (2)	-0.35779 (2)	NA (0)	0.344601 (20)
6 (PAQ8)	-0.36461 (10)	NA (0)	0.22963 (15)	NA (0)	0.319833 (7)
7 (PBQ1)	-0.27945 (10)	NA (0)	-0.09474 (13)	NA (0)	0.67369 (9)
8 (PBQ3)	-0.50945 (6)	-0.14062 (9)	0.047667 (9)	0.626158 (3)	0.810389 (5)
9 (PBQ4)	-0.24599 (13)	0.06276 (7)	0.274254 (8)	-0.03798 (1)	0.879887 (3)
10 (PBQ5)	-0.08618 (11)	NA (0)	NA (0)	NA (0)	0.142151 (21)
11 (PBQ6)	-0.02869 (18)	NA (0)	NA (0)	NA (0)	0.182401 (14)
12 (PBQ7)	-0.70858 (1)	NA (0)	NA (0)	NA (0)	0.088572 (31)
13 (PCQ1)	-0.41589 (12)	NA (0)	0.183392 (4)	NA (0)	0.393393 (16)
14 (PCQ2)	0.368727 (3)	NA (0)	NA (0)	NA (0)	0.032102 (29)
15 (PCQ3)	-0.33832 (12)	NA (0)	0.040479 (12)	NA (0)	0.701403 (8)
16 (PCQ5)	-0.03491 (12)	NA (0)	NA (0)	NA (0)	0.122805 (20)
17 (PCQ7)	-0.00415 (1)	NA (0)	-0.22311 (7)	NA (0)	0.150128 (24)
18 (PCQ8)	-0.05852 (4)	NA (0)	0.032357 (25)	NA (0)	0.487441 (3)
19 (PDQ1)	-0.63389 (6)	-0.24203 (4)	-0.16641 (4)	0.323129 (11)	0.559976 (7)
20 (PDQ2)	-0.23441 (20)	NA (0)	0.459495 (2)	NA (0)	0.580634 (10)
21 (PDQ3)	-0.32555 (6)	NA (0)	-0.17868 (11)	NA (0)	0.397061 (15)
22 (PDQ5)	-0.03522 (24)	NA (0)	0.397667 (5)	NA (0)	0.298047 (3)
23 (PDQ6)	-0.1829 (5)	NA (0)	NA (0)	NA (0)	0.109321 (27)
24 (PDQ8)	-0.34733 (13)	-0.16427 (4)	NA (0)	0.068601 (1)	0.510063 (14)

Table 2.6: Mean blind person location for the five response categories.

2.6.4 Evidence based on fairness, item by group level

A Differential Item Functioning (DIF) analysis is one that examines item function across two subgroups of the respondent sample (Wilson, 2005, p. 152). According to (Wilson, 2005, p. 152), a requirement of acceptable item design is that the items function in a similar manner across the different subgroups for participants at the same location. For the DIF analysis here, the two subgroups were determined based on whether sighted or blind participants reported having a biology-related field of study; that is, subgroup 1 consisted of the respondents who had a biology-related field of study (Sighted: N=12; Blind: N=7), and subgroup 2 consisted of participants who did not (Sighted: N=19; Blind: N=25). If the assessment tool appropriately measures comprehension of the different passages, then the items should not function differently for these two subgroups.

Figure 2.15 shows two scatter plots for item difficulty (i.e., estimate) values for the sighted and the blind participant data, respectively. The x-axis represents estimates for those with non-biology educational backgrounds, and the y-axis displays these values for participants with biology educational backgrounds. Any point that falls near the diagonal 45-degree line indicates relatively equal item functioning across the two different subgroups (non-biology versus biology background) of each participant category (sighted versus blind). The different colored dots represent three different step thresholds between construct map levels; that is, Step 1 is the transition between “Missing/Irrelevant” and “False Recall/Inference”, Step 2 is the transition between “False Recall/Inference” and “Describing”, and Step 3 is the threshold between “Describing” and “Partial Recall/Inference”. Not enough data existed to derive estimates for Step 4, the transition between “Partial Recall/Inference” to “Complete Recall/Inference”.

For sighted participant data, 18 pairs of step parameters were examined, and Figure 2.15A shows an X-Y scatter plot of item estimates per step for the two subgroups. Among the pairs examined, eight had large DIF, six had intermediate DIF, and four had negligible DIF. After calculating the 95% confidence interval for each pair, however, the results revealed that none had statistically significant DIF.

For blind participant data, 14 pairs of step parameters were examined (shown in Figure 2.15B). Of the pairs, four had large DIF, three had intermediate DIF, and seven had negligible DIF. As with the sighted participant data, the findings revealed that all 95% confidence intervals between pairs of DIF values were overlapping.

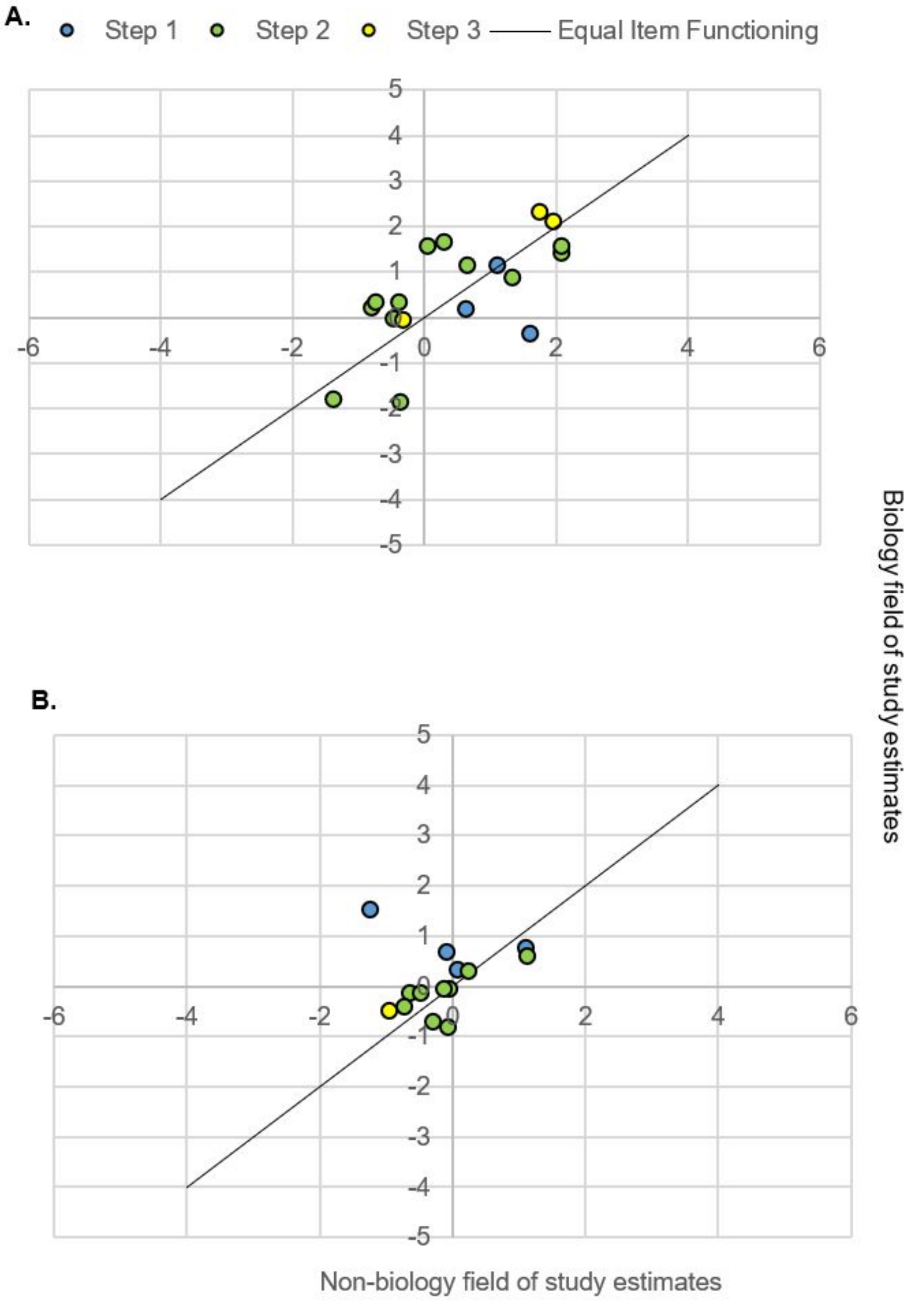


Figure 2.15: Differential item functioning analysis for (A) sighted and (B) blind participants.

2.6.5 Evidence based on response processes

All participants were asked to complete an exit survey after completing the experiment. This information serves as validity evidence, as it gives insight into the soundness of the assessment tool by comparing expectations to actual outcomes (Wilson, 2005, p. 147). Additionally, gathering response processes from an exit survey lends information about the cognitive processes participants experience while completing the assessment tool, which may provide support for the tool and/or suggestions for how to change it (Wilson, 2005, p. 161). The exit survey asked participants how satisfied they were with the voice actor and screen reader (if applicable) speech rates, how familiar they were with the passage content, and for general comments about their experience participating in the experiment. The sighted and blind participant responses are presented in Table 2.7.

		Dissatisfied	**Neutral**	**Satisfied**
Satisfaction with voice actor speech rate	Sighted	0/31	3/31	28/31
	Blind	0/32	4/32	28/32
Satisfaction with screen reader speech rate	Sighted	NA	NA	NA
	Blind	0/32	5/32	27/32
		Not Familiar	**Familiar**	**Very Familiar**
Familiarity with passage content	Sighted	7/31	19/31	14/32
	Blind	12/32	14/32	1/32

Table 2.7: Sighted and blind participants' speech rate satisfaction and familiarity.

When asked to provide general comments about the experiment, several participants noted their preferred presentation modality for comprehension during the experiment and commented on the passage content. These data are displayed in Table 2.8. "Text/braille>audio" signifies that individuals preferred text/braille to audio for comprehension, "text/braille<audio" represents greater preference for audio versus text/braille for comprehension, "text/braille = audio" indicates equal preference for the two modalities, and "difficult passages" represents individuals who felt that the passage content was too difficult to understand. Altogether, these response processes indicate that most blind and sighted participants were satisfied with the experimental parameters and passage content associated with taking the assessment tool.

	Text/Braille > Audio	**Audio > Text/Braille**	**Text/Braille = Audio**	**Difficult Passages**
Sighted	10	3	0	1
Blind	17	2	8	7

Table 2.8: Sighted and blind participants' modality preferences.

2.6.6 Evidence based on relations to other variables

Investigating the expected correlation between the instrument construct (i.e., comprehension ability) and an external variable is also a measure of validity. The external variable used here is reading fluency. Reading fluency scores were obtained from each participant by administering a subset of the Wechsler Individual Achievement Test, Third Edition (WIAT-III) before they completed the assessment tool (see Chapter 3 for further information). The WIAT-III is an isolated word reading fluency test that consists of two different parts: Word Reading and Pseudoword Decoding. Each participant's WIAT-III score was calculated by summing their scores on the Regular and Pseudoword portions of the text. The hypothesis was that participants with better reading fluency scores (i.e., higher scores) would also have better comprehension ability. For sighted participants, the Pearson correlation between oral reading fluency score and comprehension ability is 0.828 and for blind participants, it is 0.441. Indeed, correlation values between 0.36 to 0.67 represent moderate correlations and between 0.68 and 1.0 indicate strong or high correlations (Taylor, 1990). These values, therefore, indicate that there is at least a moderate correlation between the comprehension ability predicted by the assessment tool and participants' oral reading fluency scores.

2.7 Summary and conclusion

In summary, this chapter described the creation of a novel assessment tool that could be used to measure comprehension among both sighted and blind individuals. Rigorous measures were taken to ensure that the questions and responses successfully assessed comprehension ability among sighted and blind individuals. Such measures involved calibrating the test and examining its validity and reliability in assessing comprehension ability. The analyses described in this chapter suggest that this free-response comprehension assessment reasonably tests comprehension ability among sighted and blind individuals.

Having the ability to utilize the same assessment tool on two different participant groups allows for direct exploration of the cognitive mechanism underlying superior comprehension among sighted individuals; that is, for sighted individuals, is text comprehension superior to audio comprehension due to a domain-specific property of vision? Or, does this effect occur due to the physically-engaging and effortful nature of active reading compared to passive listening? Addressing these empirical questions is the focus of the present dissertation.

Chapter 3

General Methodology

3.1 Introduction

This chapter describes the methodology, design, and data analyses utilized for the experiments presented in Chapters 4-6.

3.2 Participants

3.2.1 Sighted participants

A total of 31 sighted individuals participated. The age range was 18 years to 64 years, and the median age was 22 years. Seven of the participants were male.

3.2.2 Blind participants

A total of 34 legally-blind individuals participated. The age range was 19 to 71 years, and the median age was 45 years. Thirteen of the participants were male. Furthermore, 16 participants reported having light perception, 14 had no light perception, five had perception of light projection and shadow direction, and three could perceive color or form. Twenty-four out of the 34 total participants became legally blind before the age of one year.

3.3 Materials

3.3.1 Pre-experiment survey

All participants completed a pre-experiment survey that included questions about their age, visual impairment (if applicable), field of study, education level, and amount of experience reading text/braille and listening to audio. The amount of experience was calculated by asking participants how many hours per week they engage in reading text/braille for school/work and leisure. The pre-experiment survey is provided in the Appendix and the participants' responses are given in Tables 4.2 and 5.2 in Chapters 4 and 5 of this dissertation.

3.3.2 Isolated word reading

Each participant completed the Wechsler Individual Achievement Test Third Edition (WIAT-III) isolated word reading tests: Word Reading and Pseudoword Decoding (Pearson, 2009). The Word Reading test was composed of real words (e.g., bear, ridiculous), whereas the Pseudoword Decoding test contained a series of non-words that were made to sound like actual words in the English language (e.g., snay, floit).

To make these tests accessible to blind participants, all materials were transcribed in UEB format. Both tests begin with short, simple words and progress to longer, more complicated words. Participants were scored based on their accuracy in verbally pronouncing each word. This accuracy was determined by comparing the participants' responses to a recorded answer key that is provided with the tests. The number of words correctly pronounced were summed for the Word Reading and Pseudoword Decoding tests, separately, which yielded the final scores. These reading tests were used to confirm proficiency in text and braille reading skills, and participants who had final scores below the 10th percentile for their age on the tests were excluded from the final data analyses presented in Chapters 4-6. Two of the 34 blind (participants #16 and 18) and none of the sighted participants were excluded based on these criteria.

3.3.3 Verbal working memory task

After the isolated word reading tests, participants were asked to complete a verbal working memory digit span test from the Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV Wechsler, 2008). The WAIS-IV test was administered to provide an additional variable that may help to later explain the comprehension findings. This test consisted of three different sections, during all of which the experimenter read a series of digits aloud to participants and asked them to repeat the digits in a certain order.

During the first section of the task, participants were instructed to verbally repeat the digits in the same order. The second section required participants to repeat the numbers in reverse order. For the third section, participants had to order the numbers from least to greatest. The experimenter could not repeat any of the digit sequences. Each section was completed when the participant incorrectly responded two times in a row. The score for each section corresponded to the total number of digits the participant was accurately able to remember and order.

3.3.4 Reading and listening material

Two different kinds of passages were used: practice and experimental. Seven different practice passages were used, which functioned to familiarize participants with the experimental procedure and technology used in the experiments. These passages consisted of short, literary stories acquired from the Gray's Oral Reading Fluency Test, Fifth Edition (Wiederholt & Bryant, 2012). They written at a fourth to fifth grade reading level, according to the Flesch-Kincaid Readability Test (Kincaid et al., 1975).

As mentioned in Chapter 2, for the experimental passages, permission from Pearson publishing was obtained to use four different excerpts from their high-school Prentice Hall Biology eTextbook (Miller & Levine, 2008). Scientific passages were selected for two reasons. First, blind individuals are underrepresented in STEM disciplines (Supalo et al., 2013), thereby emphasizing the need to enhance comprehension of material in STEM fields. Second, previous research has shown that comprehension ability depends on encounter type (i.e., braille versus audio) among braille readers particularly for non-fiction material (Lowenfeld, 1945; Foulke et al., 1962).

The experimental passages consisted of 130 up to 141 words, and all of them were written at a 12th-grade reading level (Kincaid et al., 1975). The topics of the passages included: zones of the ocean (Passage A), relative dating of fossils (Passage B), how viruses produce disease (Passage C), and invertebrate body symmetry (Passage D). Although the experimenters developed eight free-response questions per experimental passage (32 questions total), all questions failed to accurately assess comprehension ability among both sighted and blind participants; consequently, only six questions per passage were utilized in the final data analysis (see Chapter 2 for complete justification). For each passage, three of the questions were "literal", and the other three were "inferential". These two question types were included to investigate whether presentation modality affects comprehension differently depending on the information needed for a response (see Chapters 4 and 6).

3.3.5 Reading and listening devices

For sighted participants, passages were presented in two different formats: text and audio. For the text format, passages were printed on 8.5"x11" paper using 12-point, Times New Roman font. For the audio format, a professionally-trained male voice actor was recorded reading each passage as well as all comprehension questions in a soundproof room at the Smith-Kettlewell Eye Research Institute. These recordings were played aloud for participants during the study.

For blind individuals, the passages were presented using four different formats: (1) hard-copy braille (UEB, embossed on 11"x11.5" paper); (2) refreshable braille display (Brailliant BI 32); (3) voice actor; (4) screen reader (NVDA). The NVDA screen reader was used with the Eloquence and Vocalizer Expressive Add-on, which improved the voice clarity.

3.3.6 Post-experiment survey

At the end of the study, all participants verbally completed a post-experiment, exit survey, during which they answered questions about their participation experience, comfort level with the experimental parameters, and familiarity with passage content (see Appendix for complete document).

3.4 Procedure

3.4.1 Pre-experiment tests

All individuals gave written, informed consent to participate and be video recorded. Then, they verbally completed the pre-experiment survey. A video camera recorded all responses and was focused on the participants' hands during the entire study. Then, participants were administered the isolated word reading tests, and upon completion, they were read the study instructions aloud.

3.4.2 Practice procedures

Before reading or listening to the experimental passages, participants completed one or two different practice sessions. Two practices were used for the voice actor, braille display, and screen reader conditions. The first practice session functioned to familiarize participants with using relevant technological devices and did not require comprehension of any information in the practice passage. During this first practice, participants selected their preferred speech rate for the audio conditions. In all presentation mode conditions, participants were not timed and could make regressions, or go back in the text or audio, to clarify information. Allowing regressions and speech rate adjustments simulated realistic reading, thereby enhancing the generalizability and ecological validity of the findings. For audio conditions, regressions were made using a key press on a keyboard.

The second practice session familiarized participants with the experimental procedure. This was the only practice for the text and hard-copy braille conditions, because no technological training was required. Participants silently read or listened to the passage and verbally answered two free-response questions, which were previously recorded by the voice actor and were played aloud. Participants were also given the questions in text or braille. Once the questions were played, participants could no longer return to the passage. Participants said the word “next” to indicate that they wanted to progress to the next question.

3.4.3 Experimental procedures

Following each practice, participants completed the experimental conditions, during which they silently read or listened to each of the four scientific passages for comprehension. The experimental design was within-subjects, so all individuals completed all presentation mode conditions for their participant group. The presentation mode-experimental passage pairings and the presentation mode conditions were pseudorandomized using a Latin square design and counterbalanced across participants. Furthermore, the order of free-response questions for each passage was randomized for each participant.

Participants were instructed to notify the experimenter once they finished reading or listening to the passage. Then, the voice actor recordings of the free-response questions were played aloud. Participants were also given the questions in text or braille, and once the questions were presented, they could no longer return to the passage. They verbally responded and said “next” to progress to the next question.

3.5 Data analysis

3.5.1 Comprehension assessment scoring

The total comprehension score was calculated by summing the scores of each fitted question for a passage and dividing by the total possible points for that passage. A detailed description of the reliability and validity of the comprehension assessment and scoring procedures is provided in Chapter 2 of this dissertation.

3.5.2 Regression analysis

The comprehension scores were analyzed using linear mixed-effects regression (LMER) with the lme4 package in R 3.4.4 (R Core Team, 2018; Bates et al., 2014). This methodology was used rather than a standard, repeated-measures ANOVA due to violations of independence between measurements and varying levels of comprehension ability across participants (Baayen et al., 2008; Judd et al., 2012). That is, since participants had different comprehension abilities, their scores were not independent across presentation modes, thereby rendering each participant’s responses inter-dependent (Winter, 2013). This lack of independence was resolved by including a random effect of participant in all regression models, which functioned to account for a different “baseline” comprehension level for each participant (Winter, 2013).

All analyses started with a null model that included total score as the dependent variable and participant as the random intercept effect. Fixed effects were incrementally added to the model to investigate whether the goodness of fit was improved by using chi-square tests on the log-likelihood values to compare the different iterations of the model. Two models were only compared if they differed by one variable to ensure that the improvement or lack of improvement of fit was due to the presence of a particular variable. The final model was selected if it contained variables that significantly improved model fit (i.e., had a significant p-value from the chi-square test when it was added to the model), and it minimized the Akaike Information Criterion (AIC) value. The fixed effects tested in all models are presented in Table 3.1 on the following page. Note that the last three rows were only tested in the regression models for the blind participant data, as they are only relevant to this participant group.

Estimated marginal means, or the mean response for each presentation mode condition adjusted for the other variables in the model, were computed for all models in this study using the “emmeans” package in R 3.4.4 (Lenth et al., 2018). The 95% confidence intervals and comprehension differences between reading format conditions were computed using these values, and the differences were assessed using a Bonferroni confidence level adjustment and a Holm p-adjustment method for multiple comparisons.

Fixed Effects	**Code**
Presentation mode	1=Text (sighted); Hard-copy braille (blind) 2=Voice Actor Recordings (sighted); Braille Display (Blind) 3=Voice Actor Recordings (Blind) 4=Screen Reader (Blind)
Passage topic	A=Ocean Zones B=Fossil Relative Dating C=Viruses and Vaccines D=Invertebrate Body Symmetry
Presentation mode x Passage topic interaction	See above
Participant age	Continuous variable, obtained from pre-experiment survey
Education level	1=High School Diploma 2=Higher Education (above high school)
Primary field of study	1=Non-biology field 2=Biology-related field (above high school)
Experience using text/braille in average hours per week	Continuous variable acquired from pre-experiment survey
Experience using audio in average hours per week	Continuous variable acquired from pre-experiment survey
How long it took participants to complete each presentation mode condition (i.e., "time duration")	Calculated from video recordings in seconds
Age at which participants started to learn how to read braille (i.e., "braille age")	Continuous variable acquired from pre-experiment survey
Visual impairment onset age	1=Birth (less than one year old) 2=Other (older than one year)

Table 3.1: Fixed effects tested in the linear mixed-effects regression models.

Chapter 4

The effect of presentation mode on comprehension ability among sighted individuals

4.1 Introduction

Previous research has shown that reading information in text format supports better reading comprehension ability, on average, compared to listening to the information (e.g., Green, 1981; Varao Sousa et al., 2013). The purpose of this chapter is to replicate previous research using the novel comprehension assessment tool developed for this study. Such findings would lend additional validity to the assessment tool and allow for direct comparisons with the blind participant comprehension results, as discussed in Chapters 5 and 6.

Based on previous research, the hypothesis of this study was that presenting the nonfiction, scientific passages in text compared to voice actor format would yield significantly better performance on the comprehension assessment, thereby indicating superior comprehension ability. The methodology and analyses for this experiment are discussed in detail in Chapter 3 of this dissertation.

4.2 Results

4.2.1 Pre-experiment survey

The responses to the pre-experiment survey for all sighted participants are displayed in Table 4.2. The table displays the following information for each participant: age (in years); gender; highest level of education completed; primary field of study in school (the cells highlighted in yellow signify ones classified as biology-related); and the total number of hours on average per week each individual accessed information in text versus audio formats for work/school versus leisure purposes.

As mentioned in Chapter 3, all of these data were tested in the linear mixed-effects regression model to explore whether they were significant predictors of the comprehension assessment scores (i.e., comprehension ability). Variables were kept in the model if they significantly minimized the sum of the squared residual terms, which was reflected by comparing the log likelihood statistics between each version of the model after incrementally adding a fixed effect (Chapter 3 describes this method in detail). Indeed, including some of these data as fixed effects allowed for exploring the impact of presentation modality on comprehension while holding other variables, such as age and experience, constant.

4.2.2 Regression analysis

The fixed effects that were used in the final model for this experiment included presentation mode, passage topic, and participant age (see Table 4.2 below). As a random effect, participant was included as an intercept term. The model met all LMER assumptions. Estimated marginal means were computed for all models; the 95% confidence intervals and differences between conditions were calculated using these values, and the differences were assessed using a confidence level of 0.95 with a Bonferroni adjustment and a Holm p-adjustment method.

Fixed Effects	**Betas**	**Standard Errors**	**DF**	**Z Scores**	**P Values**
(Intercept)	0.91	0.047	47.1	19.479	0*
Voice Actor	-0.051	0.021	93	-2.422	0.015*
Passage B	-0.101	0.03	93	-3.401	0.001*
Passage C	0	0.03	93	0	1
Passage D	-0.094	0.03	93	-3.183	0.001*
Age	-0.005	0.002	31	-2.908	0.004*

Table 4.1: Statistics table for fixed effects. The text condition and Passage A were used as reference variables. *=Significant p-value at a 95% confidence level.

#	Age	Gender	Highest Level Education	Field of Study	Text Use for School/Work (hrs/wk)	Text Use for Leisure (hrs/wk)	Audio Use for School/Work (hrs/wk)	Audio Use for Leisure (hrs/wk)
1	22	Female	Bachelors	MCB/Neurobiology	20	30	0	0
2	21	Female	Undergrad	Biology	5	2	0	0
3	20	Female	High School	Bioengineering	15	10	0	10
4	20	Female	High School	Gen Ed	14	1	0	0
5	21	Female	High School	Dance	11	3	0	0
6	24	Female	Masters	EECS	28	17.5	0	3.5
7	19	Female	High School	Applied Math	2	1	0	0
8	28	Male	Masters	Law	24	3	3	0
9	22	Male	Associates	Political Science	12	0	3	1
10	21	Female	High School	Math	2-3	3-4	0 (2-3 listening to lecture webcasts)	0
11	29	Female	Associates	History	0	6	0	4
12	21	Female	Bachelors	Nutrition	1	2	0	0
13	23	Female	Bachelors	Biological Sciences	3	7	0	0
14	22	Female	Associates	Psychology/Liberal Arts	6	0	0	0
15	25	Male	High School	Nursing	1	3	0	0
16	64	Male	College, 3 years	Business/Liberal Arts	3	10	1	2
17	19	Female	College, 2 years	Economics	9	2	0	0
18	19	Male	Some college	Business	10	5	2	0
19	18	Female	High School	Computer Science	5	0	0	0
20	22	Female	Bachelors	Managerial Economics	7	0	1	0
21	49	Female	College	Fine Arts	0	12	0	0
22	18	Female	High School	Computer Science	3	1	0	0
23	18	Male	High School	Operations Research	1-2	0.25	0	0
24	32	Female	Associates	Sociology/Gender Studies	4	2	0	0
25	21	Female	High School	Integrative Biology	15	5	0	0
26	23	Female	High School	Geophysics	10	0	0	0
27	27	Female	Bachelors	Pharmacy	3	2	2	0
28	23	Female	Bachelors	Human Biology/Optometry	40	2	0	0
29	19	Female	High School	Math	28	4	0	0
30	20	Female	High School	Economics	1	2	0	0
31	24	Male	Bachelors	Fiber Polymer Science	35	5	0	0

Table 4.2: Pre-experiment survey responses from sighted participants.

4.2.3 Impact of presentation mode and passage topic on comprehension

Passage topic and age were controlled to assess the differences in total score between presentation modes. The average total score for sighted participants was significantly better with the text (mean=0.748; 95% CI: 0.705 to 0.791) compared to voice actor (mean=0.697; 95% CI: 0.654 to 0.740) format (p=0.0198; Figure 4.1). This finding agrees with previous research (e.g., Green, 1981; Varao Sousa et al., 2013).

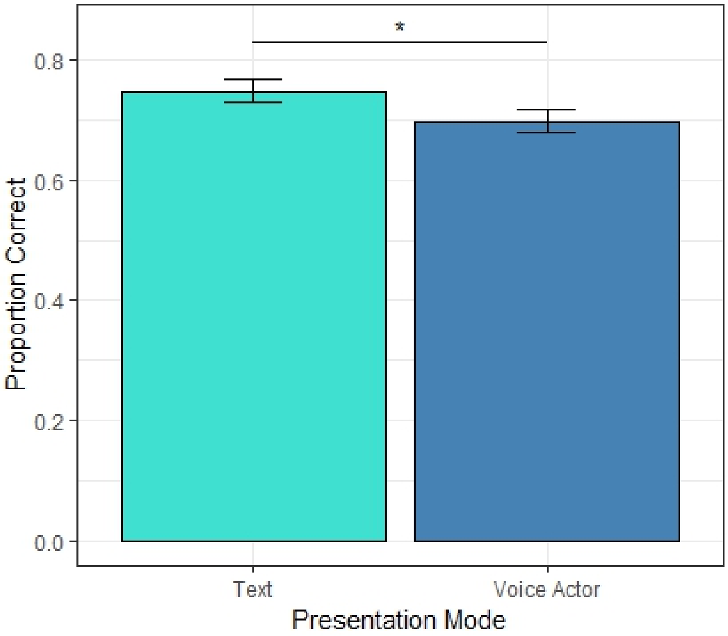


Figure 4.1: Effect of presentation mode on comprehension ability among sighted participants, controlling for participant age and passage topic. The error bars represent the standard error of the estimated marginal mean.

4.2.4 Comprehension differences between passage topics

Holding presentation mode and age constant, the differences in the proportion correct between passage topics were assessed. As shown in Figure 4.2, the average total score was significantly better for Passage A (mean=0.771; 95% CI: 0.710 to 0.832) compared to B (mean=0.671; 95% CI: 0.61 to 0.732; p=0.007) and Passage D (mean=0.677; 95% CI: 0.616 to 0.738; p=0.01). It was also better for Passage C compared to B (mean=0.771; 95% CI: 0.710 to 0.832; p=0.007) and Passage C compared to D (p=0.01). Nevertheless, there was no significant interaction between presentation mode and passage topic, meaning that for all passage topics, text supported superior comprehension compared to the audio format.

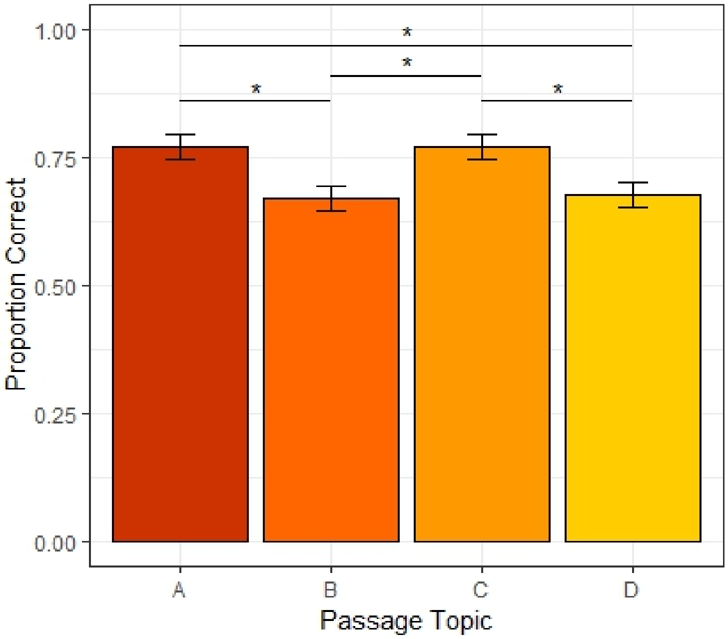


Figure 4.2: Effect of passage topic on comprehension ability among sighted participants, controlling for participant age and presentation mode. The error bars represent the standard error of the estimated marginal mean.

4.2.5 Linear relationship between significant continuous fixed effects

The fixed effect of participant age was a significant predictor of the comprehension score among sighted individuals. The linear relationship between age and total score was assessed using the Pearson correlation function in R 3.4.4. The findings revealed a negative correlation between these two variables (r=-0.29; Figure 4.3). Specifically, for the text format condition, the correlation between and regression equation for age and the total score were r=-0.398 and y=-0.006x+0.899. For the audio format condition, the correlation between and regression equation for these two variables were r=-0.2 and y=-0.003x+0.774. This negative linear relationship implies that as age increases, the average comprehension score slightly decreases; however, by looking at Figure 4.3, this effect seems driven by a couple of participants aged 49 and 64 years old, thereby complicating further interpretation of this relationship.

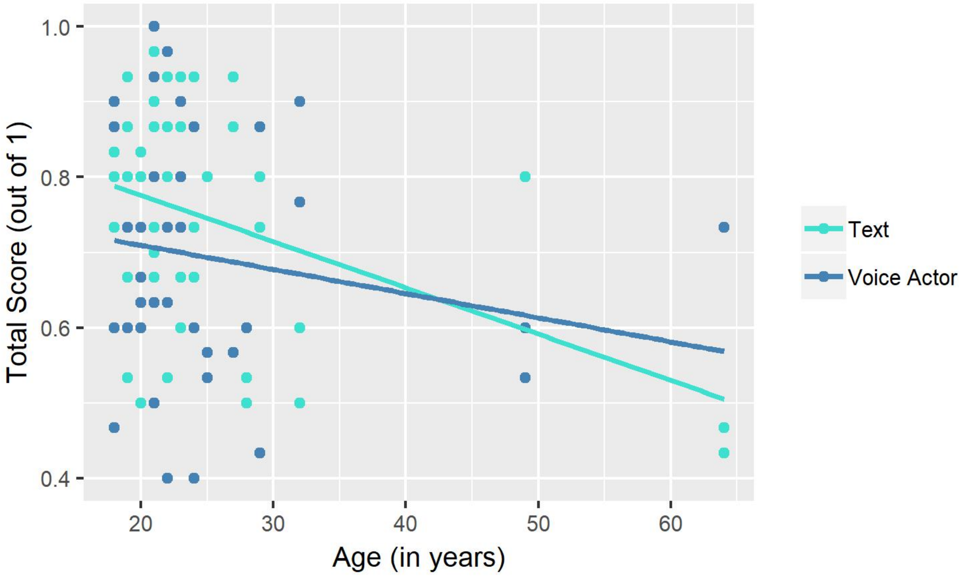


Figure 4.3: Correlation between participant age and score on the comprehension assessment, for text and audio presentation mode conditions.

4.2.6 Exploring the impact of presentation mode on time duration

The amount of time it took each participant to finish reading a passage for comprehension in each condition was calculated from the video recordings. Although time duration was not a significant fixed effect in the linear mixed-effects regression model, the results were further explored to investigate the potential trade-off between the amount of time required to read a passage for comprehension and comprehension ability. Studying this trade-off is important insofar as the results will reveal more specific educational and reading recommendations, depending on what is most important to the individual (i.e., time or comprehension).

First, outliers in time duration were removed from further analysis; data were labeled outliers if they fell outside 1.5 times the interquartile range above the upper quartile and below the lower quartile. As the estimated marginal means were used for all analyses done thus far, they were also utilized for evaluating the differences in time duration between presentation mode conditions; in other words, for the analysis presented below, time duration served as the dependent measure rather than the total comprehension score. The same fixed effects and random intercept effect used previously were included in this LMER model, and the model satisfied the assumptions of normal distribution of the residuals and homoskedasticity.

The impact of presentation mode on time duration for sighted participants is shown in Figure 4.4. The results reveal that the average amount of time it took participants to read text (mean=1.258 minutes; 95% CI: 1.147 to 1.368) was not significantly different from how long it took them to listen to the voice actor for comprehension (mean=1.223 minutes; 95% CI: 1.118 to 1.332; p=0.583).

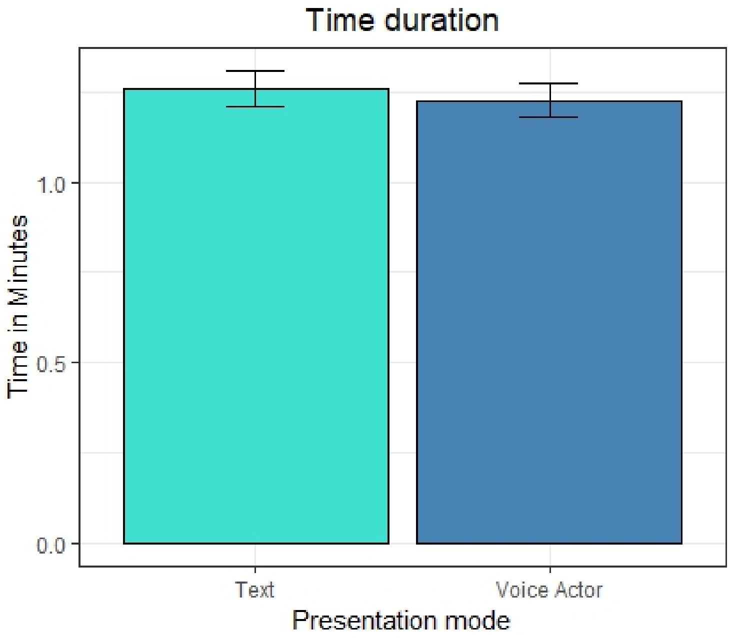


Figure 4.4: The predicted amount of time (in minutes) it took sighted participants to complete each presentation mode condition on average, averaging across passage type. The error bars represent the standard error of the estimated marginal mean.

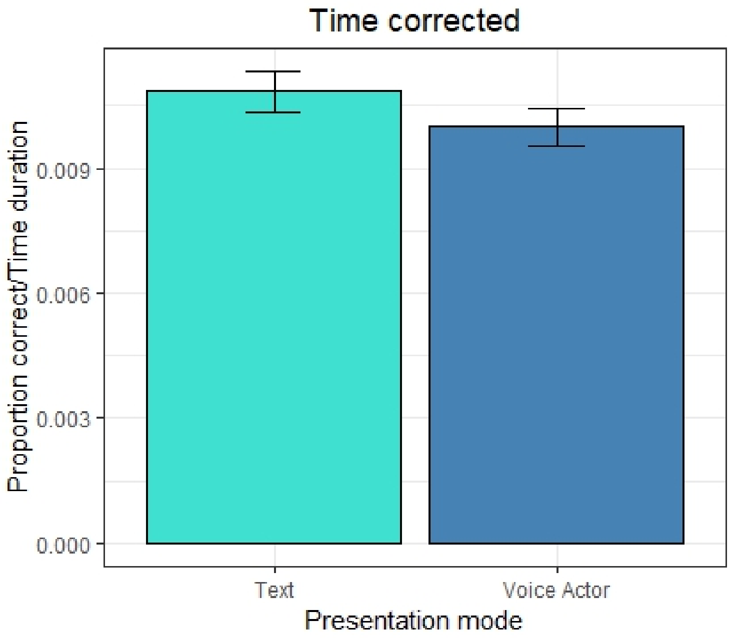


Figure 4.5: The predicted average time-corrected comprehension score for each presentation mode condition. The error bars represent the standard error of the estimated marginal mean.

The difference between the time-corrected comprehension scores for the two presentation formats was assessed. Specifically, each total score was divided by the total amount of time it took for each participant to read or listen to the passage for comprehension. After removing outliers using the same methodology described above, the estimated marginal means of the time-corrected comprehension scores for both presentation mode conditions were computed and are displayed in Figure 4.5. The findings revealed no significant difference between the average time-corrected scores for the text format (mean=0.0108; 95% CI: 0.0075 to 0.0119) and the voice actor format (mean=0.01; 95% CI: 0.0089 to 0.0105; p=0.118).

4.3 Conclusions

The findings from the experiment described in this chapter revealed that when sighted participants received the biology passages in text format, they comprehended the information more successfully than when the passages were given in auditory (voice actor) format. Additionally, there were no significant differences in the amount of time it took participants to read the passages in text compared to audio format. These results replicate previous research and provide further validity to the comprehension assessment developed for the studies described in this dissertation.

Chapter 5

Comparing the impact of presentation mode on comprehension for blind participants

5.1 Introduction

While text has been shown to support better comprehension compared to an audio format among sighted individuals, the reason for this comprehension advantage remains unknown. This finding may be a domain-specific property of the visual system; perhaps engaging the visual system itself bolsters comprehension ability. Indeed, research has shown that visual reading systematically activates a brain area in the lateral occipitotemporal sulcus, called the visual word form area (VWFA), and lesions to this area cause pure word blindness, or alexia (Dehaene & Cohen, 2011). Alternatively, the comprehension benefit for text compared to audio may exist because visual reading is more physically-engaging and effortful than passively listening. Research has shown that more difficult tasks lead to less frequent mind wandering (i.e., distractions away from the task at hand) (Smallwood et al., 2007; Forster & Lavie, 2009; Thomson et al., 2013) and deeper cognitive processing (Diemand-Yauman et al., 2011). Increasing active, physical engagement may make a task more cognitively effortful and difficult (Varao Sousa et al., 2013). If this is the case, then one would expect that blind, braille readers would experience less mind wandering and deeper cognitive processing when accessing materials in braille compared to auditory formats, as braille reading is more physically-engaging than passively listening to audio.

The present chapter assesses how blind, fluent braille readers performed on the comprehension assessment when the passages were presented in hard-copy braille compared to voice actor format. Although these individuals participated in four different presentation mode conditions (hard-copy braille, braille display, voice actor, and screen reader; see Chapter 3), only the hard-copy braille and voice actor conditions are discussed in this chapter, as these are the most analogous to the sighted participant conditions and allow for comparisons between the sighted and blind participant groups. Such comparison will explore the hypothesis that more physically-engaging and cognitively-effortful tasks promote reading comprehension, rather than being a domain-specific property of the visual system. A comparison of comprehension ability between all four presentation mode conditions among blind participants is discussed in detail in Chapter 6.

5.2 Results

5.2.1 Pre-experiment survey

The responses to the pre-experiment survey for all blind participants are displayed in Table 5.2. The table displays the following information for each participant: age (in years); gender; highest level of education completed; primary field of study in school (the cells highlighted in yellow signify ones classified as biology-related); level of vision loss; etiology of vision loss and onset of visual impairment; age at which each participant learned to read braille; and the total number of hours on average per week each individual accesses information using hard-copy braille, a refreshable braille display, a voice actor, and a screen reader for work/school versus leisure purposes.

As discussed in Chapter 3, all of these data were used in the linear mixed-effects regression model to explore whether they were significant predictors or explainers of the scores on the comprehension assessment (i.e., comprehension ability).

5.2.2 Regression analysis

The fixed effects that were used in the final model included presentation mode, participant age, and age at which the participant learned to read braille (i.e., braille age). As a random effect, participant was included as an intercept term (see Table 5.1). The model met all LMER assumptions. As discussed in Chapter 3, estimated marginal means were computed for all models; the 95% confidence intervals and differences between conditions were calculated using these values, and the differences were assessed using a confidence level of 0.95 with a Bonferroni adjustment and a Holm p-adjustment method.

Fixed Effects	Betas	Standard Errors	DF	Z Scores	P Values
(Intercept)	1.022	0.082	35.35	12.502	0*
Voice Actor	-0.085	0.037	32	-2.338	0.019*
Age	-0.005	0.001	32	-3.625	0*
Braille Age	-0.009	0.004	32	-2.504	0.012*

Table 5.1: Statistics table for fixed effects. The hard-copy braille condition was used as a reference variable. *=Significant p-value at a 95% confidence level.

Table 5.2: (below) Pre-experiment survey responses from blind participants.

#	Age	Gender	Highest Level Education	Field of Study	Level of Vision Loss	Cause of Vision Loss and Onset	Braille Age (yrs)	Braille Use for School/ Work (hrs/wk)	Braille Use for Leisure (hrs/wk)	Audio Use for School/ Work (hrs/wk)	Audio Use for Leisure (hrs/wk)	Total Screen Reader Use (yrs)	Total Braille Display Use (yrs)
1	33	Male	Some college	Electrical engineer	LP	GLC/ Birth	5	1	16	8	20	23-24	16
2	29	Female	Vocational school	Court reporting	L,SD	Optic Atrophy/ Birth	5	0	0	0	15	20	15
3	27	Female	Some college	Psychology	LP/C	GLC/Birth	6	0	0	20	10	15	UK
4	66	Female	Masters	English	NLP	ROP/Birth	5	0.5	8	0	14	21	UK
5	45	Female	Some college	Journalism/ Broadcasting	NLP	No optic nerve/Birth	4-5	0	0	10-15	7-10	32	0
6	40	Female	Bachelors	Social Sciences	NLP	RB/13 mo.	6	0	2-3	0	17	20	15
7	58	Female	Bachelors	Psychology	LP	Lead poison/18 mo.	6	2-4	8	0	40	27	UK
8	71	Male	Graduate school	Special education	LP	Optic nerve damage/ Birth	6	0	5	0	10	30	10
9	55	Male	Bachelors	Business/ Computer Information	NLP	ROP/ Birth	5	4	2	1	3	20	2
10	45	Male	Bachelors	Criminal justices	NLP	Uveitis, GLC, Cat/22 yrs	33	0	7-14	0	5-6	8	3
11	19	Male	High School	Social work	L,SD/ C/F	RP/Birth	8	3	0	0	15	7	4
12	24	Female	Vocational school	Linguistics	L,SD	LCA/Birth	3	1	0.5	4	5	6	15
13	35	Male	Masters	Vocational rehabilitation counseling	NLP	RB/2 yrs	4	21	0	40	15	31	5

14	68	Male	Bachelor of Arts	Political science/ international relations	NLP	ROP/Birth	5	30	15	25	20	34	22
15	38	Female	Associates	Health	NLP	GLC/Birth	6	5	5	0	15	31	5
17	51	Female	PhD	Computer science and music	L,SD	LCA/Birth	5	2	0	28	10	25	15
19	48	Male	Some college	Computer science	LP	Optic nerve damage/2 yrs	5	8	1	0	25	35	12
20	32	Female	Trade school	Massage therapy	NLP	ROP/Birth	5-6	0	0	0	25	20	0
21	69	Female	Masters	Rehabilitation counseling	LP	ROP/Birth	16	0	10	0	50	27-28	0
22	58	Male	Associates	General education/liberal arts	LP	Accident in OS; Cat, RD in OD/12 yrs	13	7.5	0	5	7	19	0
23	63	Female	Masters	Special education for the visually impaired	NLP	ROP/Birth	13	40	0	0	14	16	20
24	32	Male	Bachelors	Media arts and computer animation	NLP	RAC/29 yrs	30	0	0	12	116	3	2
25	56	Female	Bachelors	Languages	NLP	RB/8 mo.	4	0	0	0	30	0	0
26	25	Female	Bachelors	International relations	L,SD	LCA/Birth	13	0	1	15	10	9	10
27	23	Male	High school	Communications	NLP	TBI/4 yrs	6	0	2.5	4.5	2.5	7	7
28	24	Female	Associates	Spanish	C/F	Albinism/ Birth	14	6	4.5	16	16	18	17
29	23	Female	High school	Psychology	L,SD/ C/F	GCA/Birth	8	0	0	10	5	10	3
30	55	Male	Bachelors	Business administration/ communication	NLP	LCA/ Birth	5	6-12	5-10	10	15-20	30	20
31	61	Female	Masters	English and writing	LP	Macular scars, RP/ Birth	5.5-6	51	10	10	10	28	34

32	66	Male	Bachelors	Psychology	NLP	ROP/Birth	5	3	15	20	15	34	20
33	53	Female	Bachelors	Psychology/health and development	LP	RD/Birth	5-6	12	12	11	25	35-40	25-30
34	70	Female	Masters	Languages/Latin American literature	LP	ROP/Birth	6	14	7	14	0	0	27

The following abbreviations have the corresponding meanings:
UK=Unknown
Perception Abbreviations:
LP=Light Perception;
NLP=No light perception;
C=Color perception;
F=Form;
L,SD=Light and shadow direction;
Visual Impairment Abbreviations:
GLC=Glaucoma;
ROP=Retinopathy of prematurity;
RD=Retinal Damage;
Cat = Cataract;
RP = Retinitis pigmentosa;
LCA = Leber congenital amaurosis;
RB = Retinoblastoma;
TBI = Traumatic brain injury

5.2.3 Impact of presentation mode on comprehension

Participant age and braille age were controlled to assess the differences in total comprehension score between the hard-copy braille and voice actor conditions for the blind participants. The average total score was significantly better with the hard-copy braille format (mean=0.704; 95% CI: 0.633 to 0.775) compared to the voice actor format (mean=0.619; 95% CI: 0.547 to 0.690) format (p=0.028; Figure 5.1).

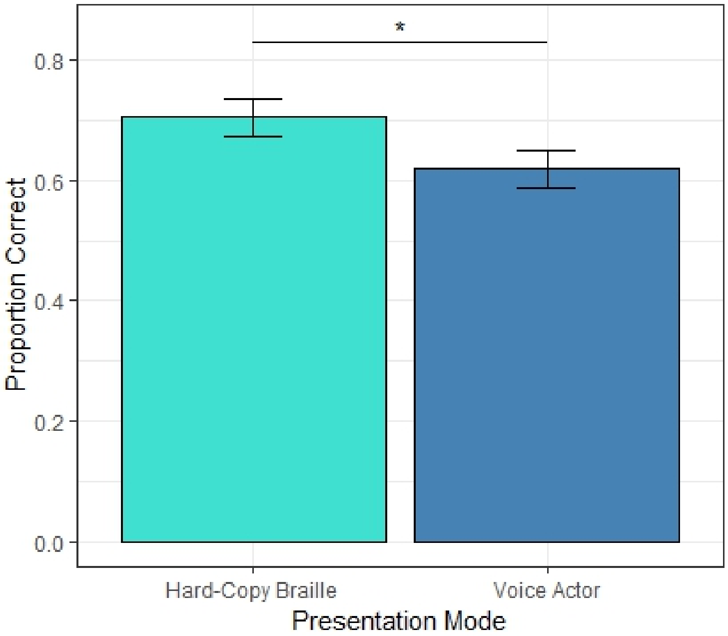


Figure 5.1: Effect of two presentation modes (hard-copy braille versus voice actor) on comprehension ability among blind, fluent braille readers. Error bars represent the standard errors of the estimated marginal means.

5.2.4 Linear relationship between significant continuous fixed effects

The fixed effects of participant age and braille age were significant predictors of the comprehension score among blind individuals. The linear relationship between participant age and total score was assessed using the Pearson correlation function in R 3.4.4. The findings revealed a negative correlation between these two variables (r=-0.40; Figure 5.2). The correlation values and regression equations were also calculated between age and the total score for the hard-copy braille condition (r=-0.366; y=-0.004x+0.908) and the voice actor condition (r=-0.453; y=-0.005x+0.868). Comparing these two values reveals that there is a slightly stronger negative correlation between these two variables for the voice actor compared to hard-copy braille condition.

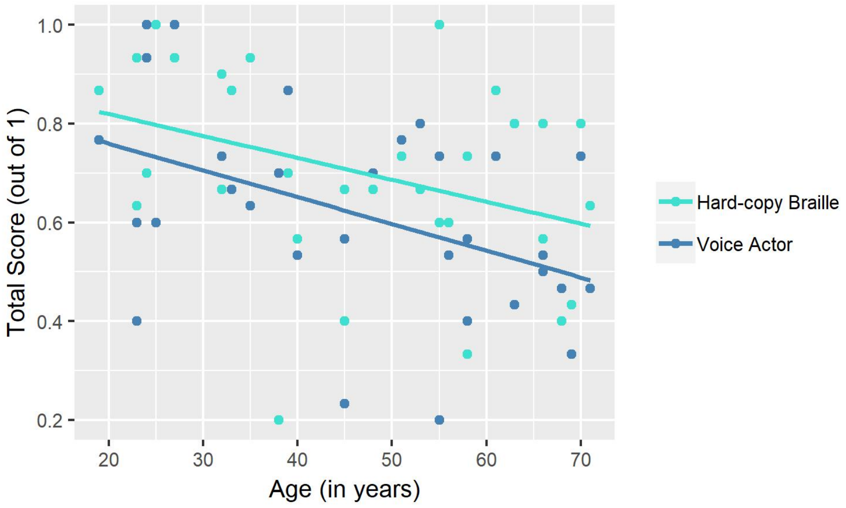


Figure 5.2: Correlation between participant age and the proportion correct on the comprehension assessment for each of the two presentation mode conditions.

The correlation between braille age and proportion correct on the comprehension assessment was also a negative value (r=-0.254). When separated by presentation mode condition, the correlation coefficients and regression equations for braille age and comprehension ability in the hard-copy braille condition were r=-0.275 and y=-0.008x+0.772. These values for the voice actor condition were r=-0.244 and y=-0.007x+0.678 (Figure 5.3).

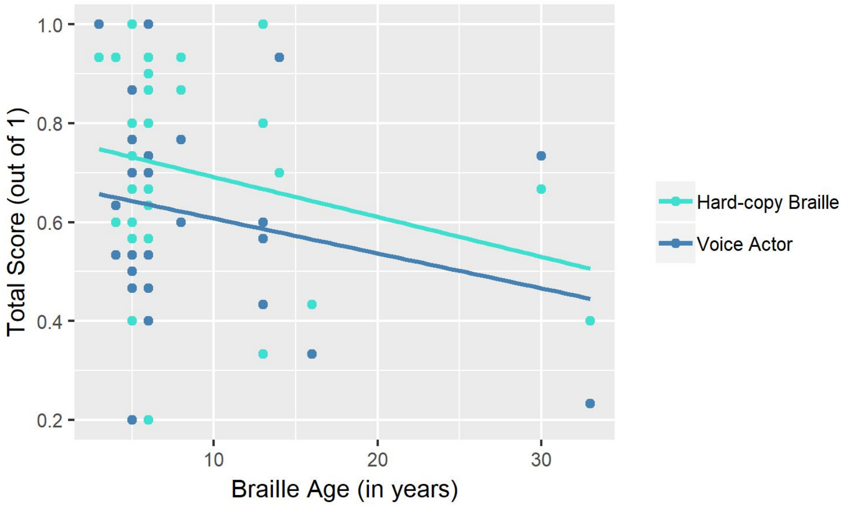


Figure 5.3: Correlation between the age during which participants learned to read braille and their score of the comprehension assessment for each of two presentation mode conditions.

Although not included in the regression model, it is insightful to assess the correlation between braille experience (i.e., the difference between the participants' age and the age at which the participants learned to read braille), as this lends a different perspective of the analyses provided above by combining the two fixed effects. The correlation between braille experience and the total comprehension score was r=-0.263. Furthermore, when separated by reading format type, the correlation and regression equation between braille experience and comprehension in the hard-copy braille condition were r=-0.225 and y=-0.002x+0.796. For the voice actor condition, these values were r=-0.314 and y=-0.003x+0.745 (Figure 5.4). The negative correlation between braille experience and comprehension ability for the voice actor condition may be expected if participants with greater braille experience spent more time using braille compared to audio formats. The negative correlation for the hard-copy braille condition, however, was unexpected. One possible explanation for this finding is that perhaps those with more braille experience were not as familiar with the UEB format used for the braille condition, as it is a relatively new format (see Chapter 1 for additional information). Another potential reason for this result is that the difference between participant age and braille age may not accurately reflect the amount of braille experience. For example, a 60-year-old individual who learned to read braille at the age of five years old would, according to this metric, have 55 years of braille experience; however, this individual may not have utilized braille for reading during this entire time, thereby contributing to the inverse relationship found between comprehension and braille experience for the hard-copy braille condition.

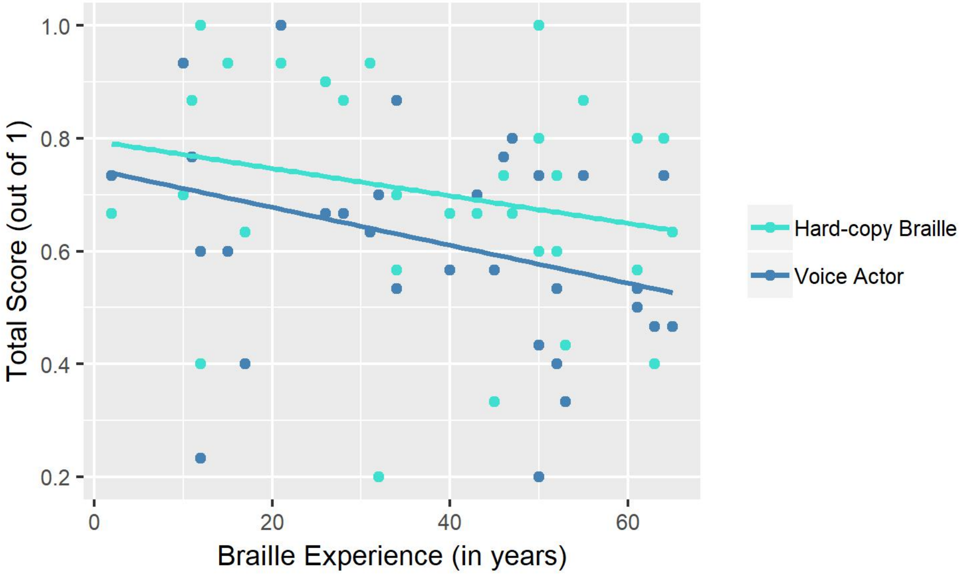


Figure 5.4: Correlation between the amount of braille reading experience and their score of the comprehension assessment for each of two presentation mode conditions.

Taken together, the negative correlations found between participant age and comprehension ability and between braille age and performance on the comprehension assessment could potentially be explained by the fact that working memory capacity and meta-cognition processes, which are involved in reading comprehension, are shown to worsen with age (Beni & Cornoldi, 1985; Carretti et al., 2007). It should be noted, however, that these negative correlation values are not likely due to any declining tactile acuity with age; previous research has shown that while the tactile acuity of sighted individuals deteriorates with age, it is retained throughout the lifespan in blindness (Legge et al., 2008).

5.2.5 Exploring the impact of presentation mode on time duration

The amount of time it took each blind participant to finish reading a passage for comprehension in the hard-copy braille and voice actor conditions was calculated from the video recordings. As in the previous chapter, although time duration was not a significant fixed effect in the linear mixed-effects regression model, the results were further explored to investigate the potential trade-off between the amount of time required to read a passage for comprehension and comprehension ability.

The same procedures described in Chapter 4 were utilized here to evaluate the differences in time duration between presentation mode conditions. The effects of the hard-copy braille and voice actor conditions on time duration for blind participants is shown in Figure 5.5. The findings reveal that the average amount of time it took participants to read hard-copy braille (mean=3.25 minutes; 95% CI: 2.72 to 3.78) is significantly longer on average than it took them to listen to the voice actor for comprehension (mean=1.57 minutes; 95% CI: 1.07 to 2.08; p<0.0001).

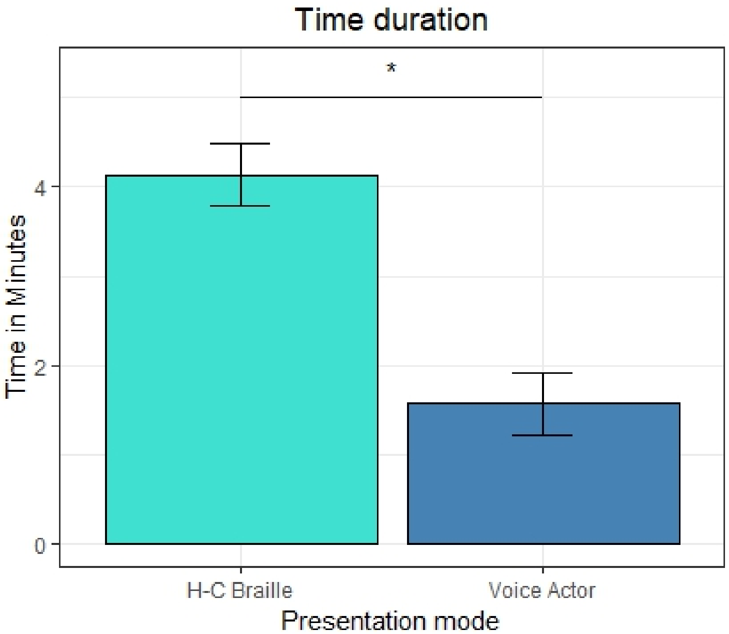


Figure 5.5: The predicted amount of time (in minutes) it took blind participants to complete each presentation mode condition on average. The error bars represent the standard error of the estimated marginal mean.

Additionally, as done in Chapter 4, the difference between the time-corrected comprehension scores for the two presentation formats (i.e., hard-copy braille and voice actor formats) was evaluated. After removing outliers, the estimated marginal means of the time-corrected comprehension scores for both presentation mode conditions were calculated and are shown in Figure 5.6. The results revealed that, after correcting for time, comprehension ability was significantly worse when reading using hard-copy braille (mean=0.0046; 95% CI: 0.003 to 0.0063) compared to listening to the voice actor (mean=0.0084; 95% CI: 0.0068 to 0.01; p<0.0001). This finding does not necessarily support the use of audio formats over braille formats for reading, as this decision depends on the situation and the reader's priorities. For instance, if an individual does not have a lot of time to study for an exam, according to these results, it would benefit comprehension more if the material is accessed using an audio rather than braille format. If, on the other hand, the individual has a longer amount of time to prepare for the exam, he or she may want to consider using a braille format, as this mode yields better comprehension on average when allowing individuals to spend as much time studying the passages as desired or needed. Indeed, additional justifications and applications for this finding are explored in Chapter 7 of this dissertation.

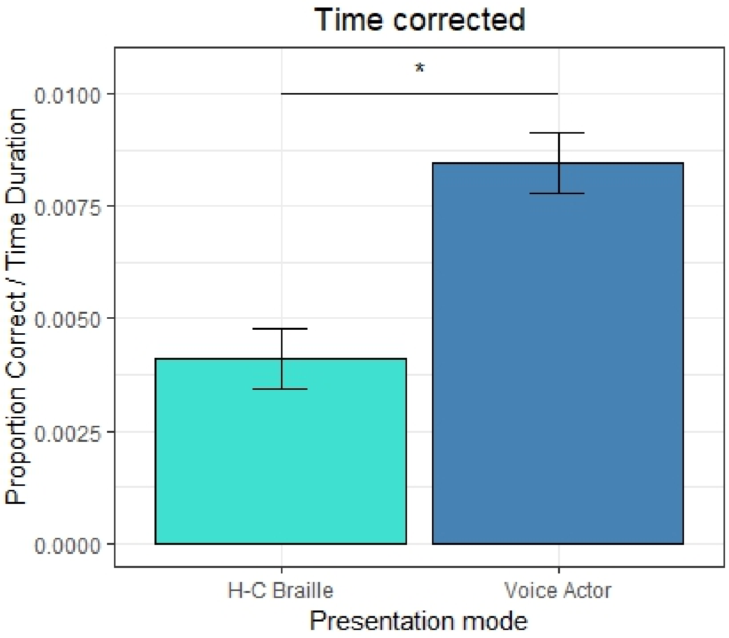


Figure 5.6: The predicted average time-correcteded comprehension score for each presentation mode condition. The error bars represent the standards error of the estimated marginal mean.

5.2.6 Comparing the effect of presentation mode on comprehension between blind and sighted individuals

The purpose of this chapter was to compare the impact of two analogous pairs of presentation mode conditions – text/hard-copy braille versus voice actor – on reading comprehension to elucidate the neural underpinnings of this cognitive process. As average comprehension was significantly better in text and hard-copy braille conditions compared to the voice actor condition (Figure 5.7), these findings provide potential support for the theory that tasks requiring greater physical engagement promote reading comprehension and imply that the visual system itself is not necessarily a fundamental component of this cognitive process. It should be noted, however, that factors other than increased physical activity may contribute to the comprehension advantage when using text and hard-copy braille versus listening to a voice actor. These alternative factors will be identified and explained in the Discussion section of this dissertation.

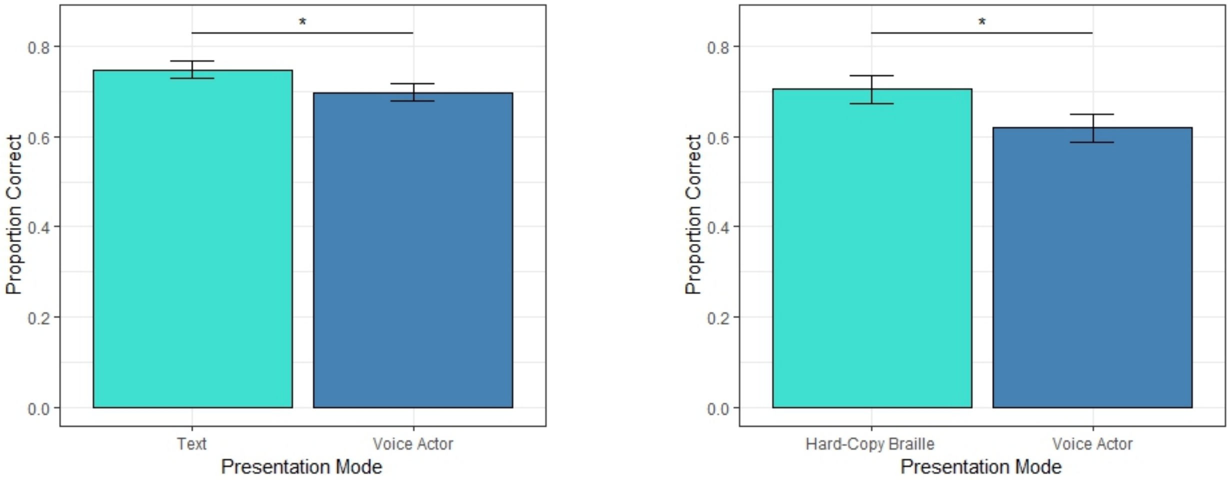


Figure 5.7: Comparing how presentation mode affects comprehension between sighted (left) and blind (right) individuals.

Furthermore, it is interesting to consider the differences in comprehension ability between the text versus hard-copy braille conditions and the two voice actor conditions across the sighted and blind participant groups. When accessing the passages using text or hard-copy braille formats, the average total score among sighted participants was 74.8% and among blind participants was 70.4%. When listening to the voice actor reading the passages aloud, the average total score among sighted participants was 69.7% and among blind participants was 61.9%. The overall poorer comprehension ability among blind compared to sighted participants may potentially be due to differences in the extent to which each passage required participants to form mental representations that were highly dependent on visual imagery and understanding. Indeed, reading materials used for many science disciplines, such as chemistry and biology, tend to cater to the visual system when explaining abstract topics, such as molecular structures and biological cycles. One of the passage topics used for the studies in this dissertation reviewed fossils and different rock layers, which may have been more easily comprehended by individuals who could form visual representations of this information. Consequently, for some of the passages used in the present studies, blind participants may have been placed at a disadvantage. This explanation could also potentially justify the surprising result that the blind participants had worse auditory comprehension compared to the sighted participants, as previous research suggests that blind individuals encode auditory information more efficiently than sighted people

(Roder et al., 2001).

5.3 Conclusions

The results in this chapter revealed that blind, fluent braille readers comprehended the nonfiction, scientific passages better on average when they read the material using hard-copy braille compared to listening a voice actor read it when not under a time constraint. With limited time, however, the results suggest that blind individuals should access reading materials using auditory rather than braille format to enhance comprehension ability. Furthermore, the comparison of comprehension results between blind and sighted participants provides potential support for the hypothesis that better comprehension ability is supported by more physically-engaging tasks. Implications of these results will be discussed in the Discussion section in Chapter 7 of this dissertation.

Chapter 6

The impact of widely-used assistive technology on comprehension ability among fluent braille readers

6.1 Introduction

The question of how presentation mode impacts comprehension ability among blind individuals has been explored by previous research (Foulke et al., 1962; Lowenfeld, 1945; Nolan, 1963); however, advances in technology have occurred since these studies were conducted. Today, many blind individuals utilize novel assistive technologies to access reading materials. Among the widely- and currently-used assistive technologies for reading include refreshable braille displays and screen readers (see Introduction for a complete description of each of these devices). The present study, therefore, incorporates these technologies to assess how they may impact comprehension of the four high school biology passages among blind, braille readers. The analyses presented in this chapter investigate the potential differences in comprehension ability that exist when these nonfiction passages are read using: (1) Embossed, hard-copy braille ("H-C Braille"), (2) A refreshable braille display ("B Display"), (3) Human voice actor recordings, and (4) A screen reader. The findings from these analyses will help clarify which presentation modalities yield better average comprehension of scientific passages. The educational implications of such results will be considered further in the following chapter.

6.2 Results

6.2.1 Regression analysis

Because all four presentation mode conditions were tested in the present LMER model, the significant fixed effects differed from what was reported in Chapter 5, when only two of the presentation modes were tested. In this model, the fixed effects that were significant and used in the final model included presentation mode, participant age, passage topic, braille age, and the interaction between presentation mode and passage topic (see Table 6.1). As a random effect, participant was included as an intercept term. The model met all LMER assumptions.

Fixed Effects	**Betas**	**Standard Errors**	**DF**	**Z Scores**	**P Values**
(Intercept)	1.043	0.077	48.11	13.583	0*
Braille Display	-0.021	0.076	100.77	-0.276	0.783
Voice Actor	-0.177	0.069	103.75	-2.578	0.01*
Screen Reader	-0.184	0.07	99.63	-2.632	0.008*
Age	-0.005	0.001	32	-4.215	0*
Passage B	0.024	0.076	100.77	0.311	0.756
Passage C	-0.13	0.07	99.63	-1.858	0.063
Passage D	-0.202	0.069	103.75	-2.942	0.003*
Braille Age	-0.006	0.003	32	-2.068	0.039*
B Display x Passage B	-0.224	0.118	70.83	-1.896	0.058
V Actor x Passage B	-0.017	0.107	80.33	-0.155	0.877
S Reader x Passage B	-0.029	0.107	97.32	-0.266	0.79
B Display x Passage C	0.091	0.107	97.32	0.846	0.398
V Actor x Passage C	0.231	0.105	102.22	2.201	0.028*
B Display x Passage C	0.261	0.113	56.71	2.305	0.021*
B Display x Passage D	0.183	0.107	80.33	1.701	0.089
V Actor x Passage D	0.169	0.108	73.92	1.569	0.117
S Reader x Passage D	0.11	0.105	102.22	1.048	0.295

Table 6.1: Statistics table for fixed effects. The hard-copy braille condition and Passage A were used as reference variables. *=Significant p-value at a 95% confidence level.

As with all of the previous analyses, differences in comprehension were assessed utilizing the estimated marginal mean values for the presentation mode conditions. These differences in comprehension between the presentation modes were investigated using a confidence level of 0.95 with a Bonferroni adjustment and a Holm p-adjustment method.

6.2.2 Impact of technology on comprehension

Participant age, braille age, and passage topic were held constant to explore the differences in comprehension ability between the four different presentation mode conditions (see Figure 6.1). The average total score was significantly better with hard-copy braille (mean=0.706, 95% CI: 0.634 to 0.777) compared to the screen reader (mean=0.607; 95% CI: 0.535 to 0.679; p=0.0227) and with the braille display (mean=0.697; 95% CI: 0.625 to 0769) compared to the screen reader format (p=0.04). Interestingly, in contrast to what was found in Chapter 5, the hard-copy braille condition no longer produced better comprehension than the voice actor condition (mean=0.625; 95% CI: 0.553 to 0.696, p=0.067). This discrepancy in findings is likely due to the novel, significant interaction that exists between presentation mode and passage topic, which signifies that the effect of presentation mode on comprehension ability depends on the passage content. This unexpected, yet interesting, interaction will be discussed further in the following chapter of this dissertation.

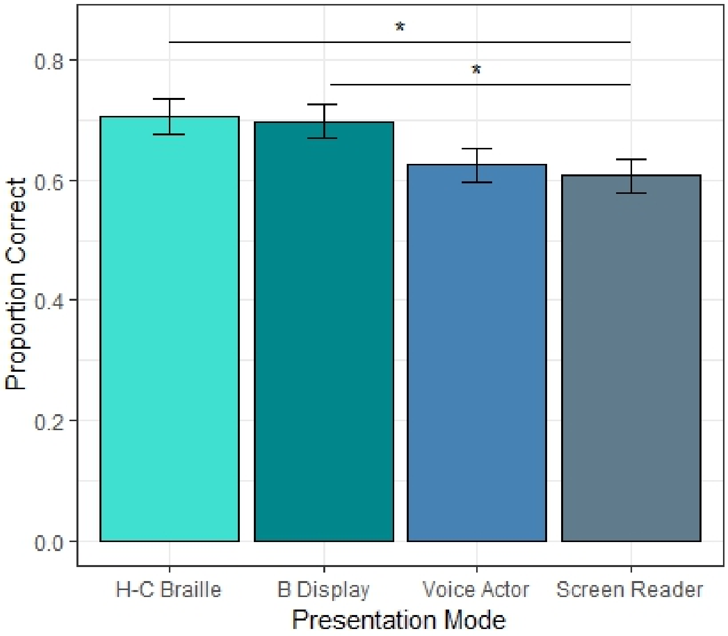


Figure 6.1: Effect of all four presentation modes on comprehension ability among blind, fluent braille readers. Error bars represent the standard errors of the estimated marginal means.

6.2.3 Comprehension differences between passage topics

Holding presentation mode, braille age, and participant age constant, the differences in the proportion correct between passage topics were assessed. As shown in Figure 6.2, the results revealed that the average total score was significantly better for Passage C (mean=0.703; 95% CI: 0.631 to 0.774) compared to D (mean=0.601; 95% CI: 0.529 to 0.672; p=0.017). Recall that Passage C was about how viruses produce disease, and Passage D was about invertebrate body symmetry. A possibility is that Passage D catered to sighted individuals, as it may have been better understood by individuals who were able to form visual representations of the symmetrical geometry. In fact, research suggests that there exist differences in how blind and sighted individuals perceive symmetry, and that the perceptual salience of symmetry in the vertical dimension is visually-based (Cattaneo et al., 2010). Passage C, on the other hand, addressed a topic (i.e., viruses and disease) that is discussed frequently in the media, making it well-known to many individuals. This aspect may have led to highest average comprehension of this passage compared to the other three.

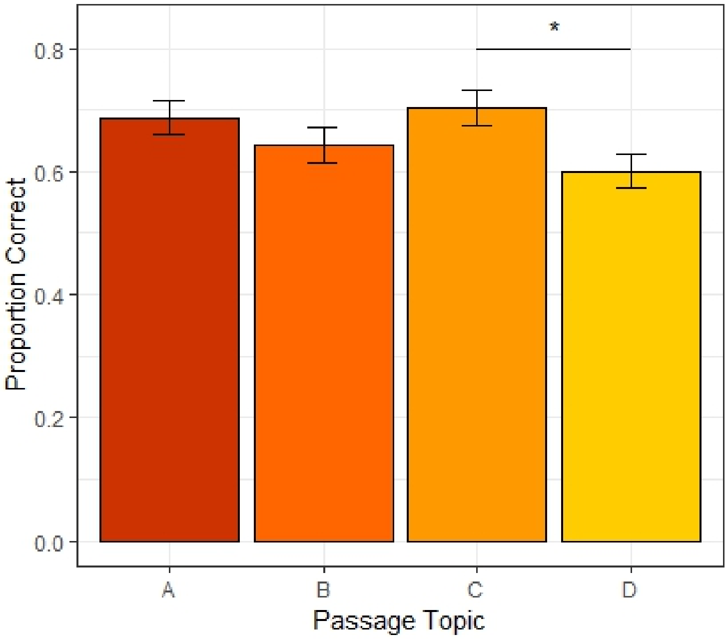


Figure 6.2: Effect of passage topic on comprehension ability. Error bars are the standard errors of the estimated marginal means.

6.2.4 Interaction between passage topic and presentation

To further explore the unexpected, significant interaction between passage topic and presentation modality, the trends in average comprehension ability between presentation modes for each passage topic are shown in Figure 6.3. As this significant interaction was not anticipated, sufficient data do not exist in each presentation mode-passage topic pairing to conduct statistical comparisons between the conditions; however, it is insightful to assess the trends insofar as this assessment may reveal information about how particular reading topics may be better understood when using a specific presentation modality.

As displayed in Figure 6.3, average comprehension of Passages A, B, and D was better with one or both braille formats compared to the auditory formats. It is interesting to consider, however, that the braille format(s) that yielded superior comprehension differed between these passages. While this finding may have been due to the insufficient sample size in each of these conditions, this trend may imply that the presentation mode most effective for maximizing comprehension ability depends on the passage content. This possibility will be discussed further in the following chapter. Additionally, the trend in average comprehension for Passage C, which discussed how viruses produce disease, revealed a ceiling effect across all presentation mode conditions. As mentioned in the section above, this result may be because this topic is frequently talked about in the media, thereby facilitating the opportunity to learn about it.

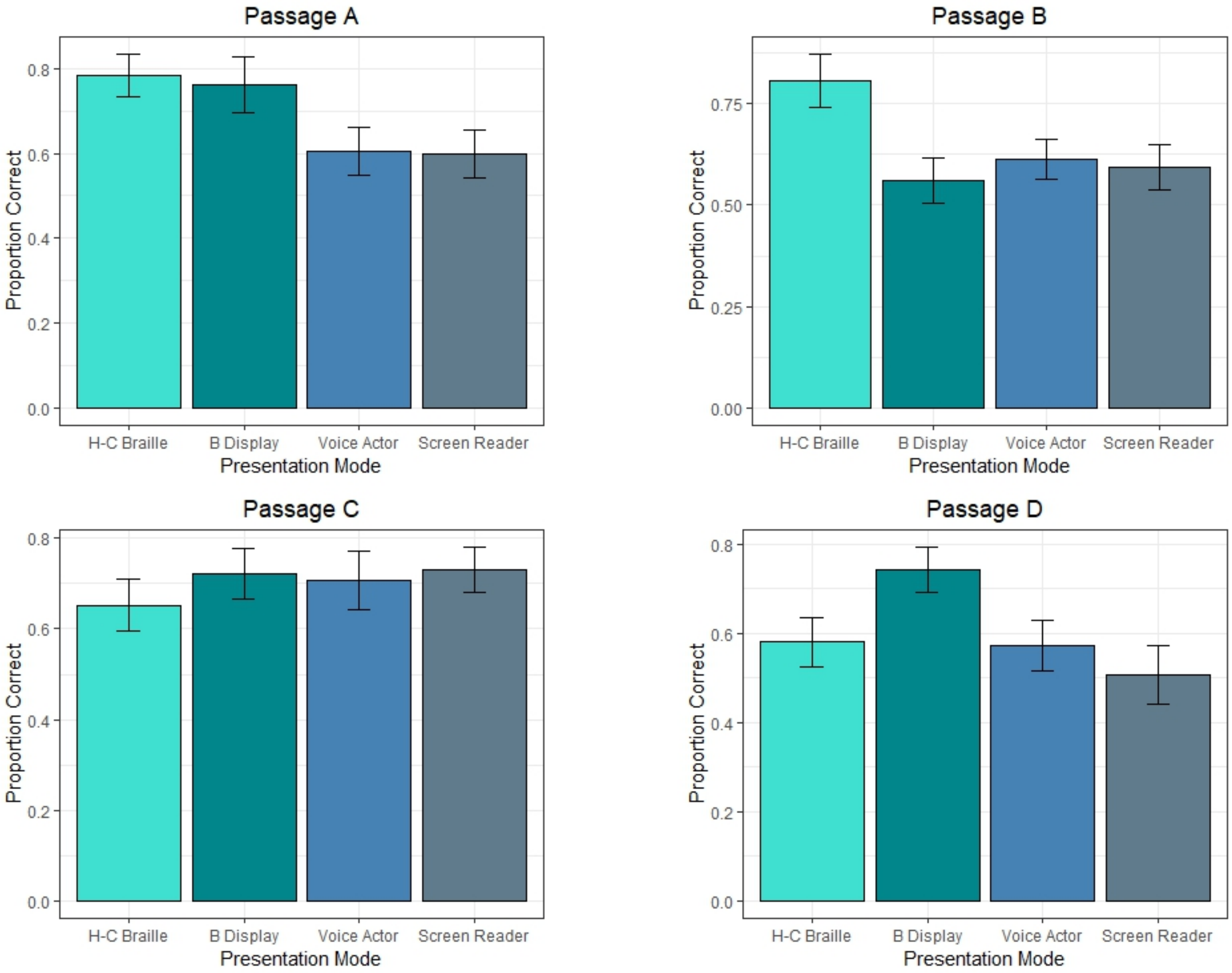


Figure 6.3: Interaction between passage topic and presentation mode. Error bars are standard errors of estimated marginal means.

6.2.5 Linear relationship between significant, continuous fixed effects

The fixed effects of participant age and braille age were significant predictors of the comprehension score when considering all four presentation mode conditions among blind individuals. The linear relationship between participant age and total score was assessed using the Pearson correlation function in R 3.4.4. The findings revealed a negative correlation between these two variables (r=-0.343, Figure 6.4). The correlation value and regression equation were also calculated between age and the total score for the hard-copy braille condition (r=-0.366; y=-0.004x+0.908), the braille display condition (r=-0.22; y=-0.002x+0.798), the voice actor condition (r=-0.453; y=-0.005x+0.868), and the screen reader condition (r=-0.357; y=-0.003x+0.768). Comparing these four values reveals that there is a slightly stronger negative correlation between these two variables for the voice actor condition compared to other conditions.

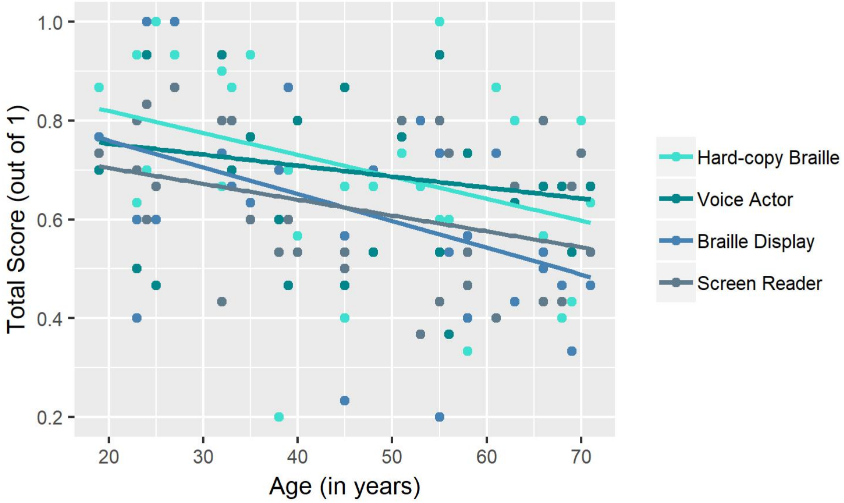


Figure 6.4: Correlation between participant age and proportion correct on the comprehension assessment for each of the four presentation mode conditions.

The correlation between braille age and proportion correct on the comprehension assessment was also a negative value (r=-0.15). When separated by presentation mode condition, the correlation between and regression equation for braille age and comprehension ability for the hard-copy braille condition were r=-0.275 and y=-0.008x+0.772, for the braille display condition were r=-0.073 and y=-0.002x+0.711, for the voice actor condition were r=-0.244 and y=-0.007x+0.678, and for the screen reader condition were r=0.037 and y=0.0008x+0.614 (Figure 6.5).

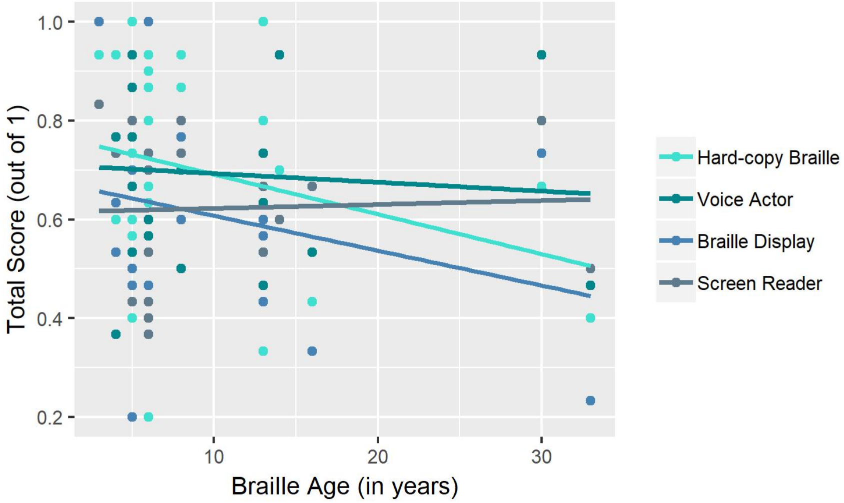


Figure 6.5: Correlation between braille age and proportion correct on the comprehension assessment for each of the four presentation mode conditions.

Interestingly, the correlation between braille age and comprehension ability was only positive for the screen reader condition compared to all other conditions; however, this positive value was very close to zero, implying a very weak relationship between these two variables for the screen reader condition. Regardless, the positive sign of the correlation coefficient signifies that the later one learns braille in life, the better their comprehension ability is when using screen reader software. This finding may be expected, as individuals who learn braille later in life may have more experience with screen reading software, since it is easier to learn and use than braille.

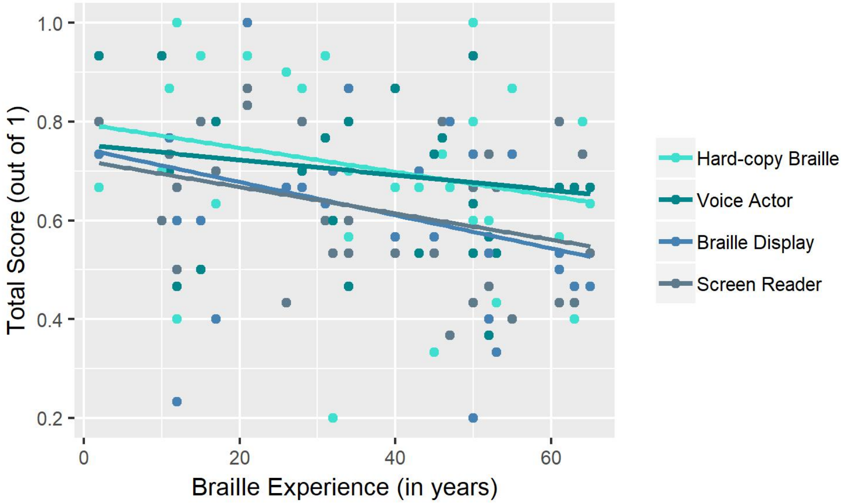


Figure 6.6: Correlation between braille experience and proportion correct on the comprehension assessment for each of the four presentation mode conditions.

To further clarify the relationship between the above fixed effects and comprehension ability, participant and braille ages were combined into the parameter "braille experience," which was calculated by taking the difference between the two effects, as was done in Chapter 5. The overall correlation between braille experience and comprehension was -0.251; but, when separated by presentation mode condition, the correlation values and regression equations were r=-0.225 and y=-0.002x+0.796 for hard-copy braille, r=-0.169 and y=-0.002x+0.753 for the braille display, r=-0.315 and y=-0.003x+0.745 for the voice actor, and r=-0.315 and y=-0.003x+0.745 for the screen reader (see Figure 6.6).

As in Chapter 5, it is somewhat surprising that there exists an inverse relationship between braille experience and comprehension score for the two braille presentation mode conditions, as one may expect that individuals with more years of braille experience individuals have should perform better on the comprehension assessment when using braille. The negative correlation value could potentially be explained using the same reasoning provided in Chapter 5: perhaps individuals with more braille experience are less familiar with the recently-established UEB braille format utilized for the braille materials in these studies. Furthermore, taking the difference between participant age and braille age may not be an accurate measure of braille experience, as individuals may not utilize braille as a reading format for that entire time.

Additionally, taken together, these statistics reveal a stronger negative correlation for the audio conditions (voice actor and screen reader) than the braille (hard-copy braille and braille display). This finding signifies that, with more years of braille experience, auditory comprehension gets worse than braille comprehension, which could be due to the greater proportion of time spent on using braille compared to audio formats for reading.

6.2.6 Exploring the impact of presentation mode on time duration

The amount of time (in minutes) that each blind participant spent on reading a passage for comprehension in the four different presentation mode conditions was calculated from the video recordings. The same procedures described in the previous chapters were utilized here to evaluate these differences, after removing outliers. These outliers were removed if they fell outside 1.5 times the interquartile range above the upper quartile and below the lower quartile. The effects of the four presentation mode conditions on time duration for fluent braille readers is shown in Figure 6.7. The findings show that the time duration was significantly longer in the refreshable braille display condition (mean=4.18 minutes; 95% CI: 3.525 to 4.827) compared to the voice actor (mean=1.59 minutes; 95% CI: 0.966 to 2.213; p<0.0001), the screen reader (mean=1.748; 95% CI: 1.125 to 2.371; p<0.0001), and the hard-copy braille (mean=3.258; 95% CI: 2.606 to 3.909; p=0.01) conditions. Additionally, the time duration was significantly slower when reading using hard-copy braille compared to listening to a voice actor (p<0.0001) and a screen reader (p<0.0001).

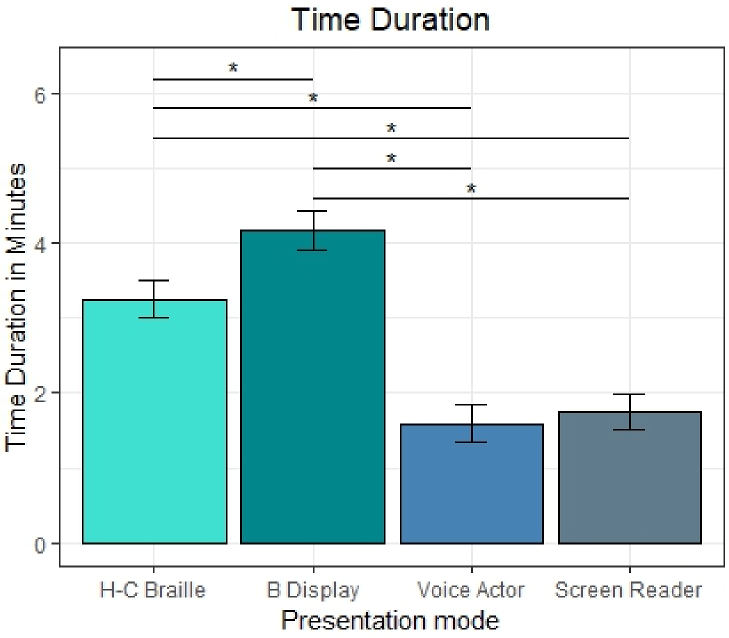


Figure 6.7: The predicted amount of time (in minutes) it took blind participants to complete each presentation mode condition on average. The error bars represent the standard error of the estimated marginal mean.

The differences between the time-corrected comprehension scores for the four different presentation formats (i.e., hard-copy braille, refreshable braille display, voice actor, and screen reader) were evaluated. After removing outliers, the estimated marginal means of the time-corrected comprehension scores for all presentation mode conditions were calculated and are shown in Figure 6.8. The results revealed that, after correcting for time, comprehension ability was significantly better when listening to a voice actor (mean=0.0083; 95% CI: 0.0061 to 0.0106) compared to reading using hard-copy braille (mean=0.0043; 95% CI: 0.0021 to 0.0065; p=0.0012) and using a refreshable braille display (mean=0.0033; 95% CI: 0.0011 to 0.0054; p=0.0001). Additionally, after correcting for time, listening to the screen reader (mean=0.009; 95% CI: 0.0068 to 0.0112) led to significantly better comprehension ability compared to when using hard-copy braille (p=0.0002) and a braille display (p<0.0001).

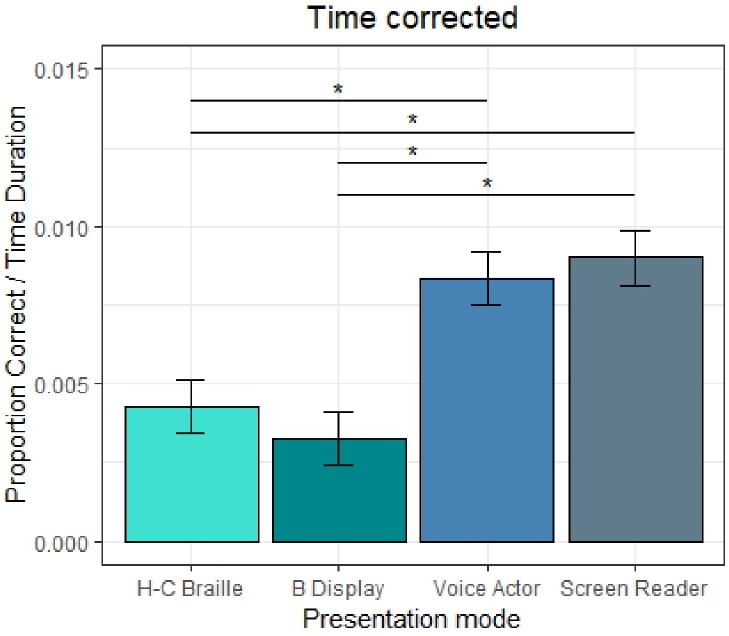


Figure 6.8: The predicted average time-corrected comprehension score for each presentation mode condition. The error bars represent the standards error of the estimated marginal mean.

6.3 Conclusions

The findings presented in this Chapter revealed that blind individuals comprehended scientific passages significantly better on average when they were presented with hard-copy braille or a refreshable braille display compared to with a screen reader. This result may be due to the poor diction and pronunciation offered by synthesized speech relative to a human voice. Notably, however, after factoring out the amount of time spent reading in each presentation mode condition, the two auditory conditions (i.e., voice actor and screen reader) yielded significantly better comprehension than the two braille conditions (i.e., embossed braille and braille display).

Consequently, as in the previous chapters, there exists a trade-off between time duration and comprehension benefit when considering the way reading materials are presented. Indeed, the greater amount of time spent when using braille compared to audio for reading may be due to more frequent regressions made in the braille relative to the audio presentation mode conditions; this hypothesis will be further explored in Chapter 7.

The analyses discussed here also showed that the impact of presentation mode on comprehension may depend on passage content. This interaction between presentation format and passage topic could potentially be explained by varying demands placed on creating spatial representations, some of which may cater to visual processing. This possibility will also be discussed further in the following chapter.

Chapter 7

General discussion

7.1 Cognitive theories for reading comprehension

7.1.1 Physical engagement and cognitive effort

One of the aims of this dissertation was to elucidate the cognitive mechanism involved in supporting reading comprehension. Indeed, previous research has shown that reading text significantly decreases mind wandering frequency and increases memory performance compared to passively listening to reading materials (Varao Sousa et al., 2013). This effect may occur due to the increased difficulty (Smallwood et al., 2007; Forster & Lavie, 2009; Thomson et al., 2013) and/or physical engagement (Clark, 1999; Wilson, 2002; Kingstone et al., 2008) associated with reading text aloud or silently compared to listening. Alternatively, reading text may promote comprehension compared to listening by engaging the visual system, signifying that visual processing is an essential mechanism for this construct. This dissertation addressed these alternative theories by comparing how presentation modality impacts comprehension ability between blind and sighted individuals.

The findings reported in Chapter 5 potentially substantiate the claim that more physically-engaging, and potentially more effortful, tasks support reading comprehension ability in blind individuals. Fluent braille readers comprehended the nonfiction, scientific passages better on average when they read the material using hard-copy braille compared to listening to a voice actor read it. Similarly, as reported in Chapter 4, the average total comprehension score for sighted participants was significantly better with text compared to voice actor. While these findings appear to provide support for the hypothesis that better comprehension ability is improved by more physically-engaging tasks, the results presented in Chapter 6 challenge this theory. That is, in Chapter 6, comprehension performance with both braille conditions (hard-copy braille and refreshable braille display) was not significantly better than with both audio conditions (voice actor and screen reader); although there was a significant difference in comprehension between the hard-copy braille and voice actor presentation modes in Chapter 5, this difference was no longer significant when analyzed with all four presentation modes in Chapter 6. As discussed in Chapter 6, this lack of significance may be due to the interaction between presentation topic and mode. Unfortunately, there did not exist sufficient data to statistically unpack and analyze this interaction; however, the trends and implications will be discussed later in this section.

The lack of significance was further explored by calculating the effect sizes for the comprehension differences between the voice actor and braille conditions utilizing a formula for Cohen's d for mixed-effects models (Brysbaert & Stevens, 2018). A low effect size (d=0.2; Cohen, 1988) would provide support for the fact that there may not actually exist a comprehension difference between the conditions, as is suggested by the p-value. A higher effect size (d=0.8; Cohen, 1988) would imply that that, although the p-value was not significant at a 95% confidence level, there may have actually existed a difference between the braille and voice actor conditions; however, it may have been masked by an extraneous, uncontrolled variable. Additionally, the number of participants tested may not have been sufficient for the amount of precision needed to make these differences statistically significant.

The observed effect sizes between the hard-copy braille and voice actor conditions was d=0.561 and between the braille display and voice actor conditions was d=0.501. These are medium effect sizes (Cohen, 1988) and may suggest that acquiring a larger sample size may have revealed a statistically significant comprehension difference between the voice actor and two braille conditions. It should be noted, however, that the sample sizes used for the studies presented in this dissertation were determined a-priori in order to avoid p-hacking. This calculation was done using G*Power (Faul et al., 2007, 2009), specifying a power of 0.8 and a repeated-measures, within-factors design. Also, a linear mixed-effects regression analysis was employed, rather than a standard repeated-measures ANOVA, to control for as many potentially confounding variables as possible. Nevertheless, the results are not purely the impact of presentation mode on comprehension, as the interaction between passage topic and presentation mode is significant.

Although, taken together, the findings presented in Chapters 5 and 6 do not provide strong or clear support for a particular comprehension mechanism, they do provide some insights. While the findings potentially corroborate the theory that active forms of reading yield better comprehension, they suggest that it may not necessarily be the case that more cognitively effortful and difficult tasks facilitate memory performance. To elaborate, one may argue that listening to a screen reader is much more difficult than reading embossed braille, due to the different keyboard combinations required to navigate through reading materials and the difference in tone, pronunciation, and diction offered by synthesized speech. If more difficult tasks support comprehension, one might expect comprehension ability to be better when using a screen reader compared to a voice actor, for instance. This finding, however, was not discovered in the studies presented here. A possibility is that the relationship between task difficulty and comprehension or memory is represented by an inverted U-shaped curve, like what has been discovered with the impact of stress on performance (Yerkes & Dodson, 1908). Perhaps the screen reader exceeded a certain amount of difficulty that would have been beneficial for comprehension. Although research has shown that greater depth of processing, which is typically associated with more difficult tasks, promotes memory performance (e.g., Craik & Tulving, 1975), these tasks may provide a moderate amount of difficulty.

7.1.2 Reading regressions

Several studies reported better average comprehension when information was read rather than listened to (e.g., Green, 1981; Varao Sousa et al., 2013). While Varao Sousa et al. (2013) contend that individuals may remember reading materials better with more active versus passive encounters, they do not consider reading regression frequency when interpreting this effect. In fact, studies that have explored differences in comprehension between text and auditory modalities have not additionally assessed regressions, nor have they reported allowing regressions in the auditory conditions (e.g., Rogowsky et al., 2016; Green, 1981). Text and braille formats may support better comprehension ability on average because the Euclidean geometry offered by paper facilitates making regressions, whereas utilizing a keyboard or keypad to regress in audio formats requires a certain amount of guessing about where in the audio the information was presented that the reader wants to clarify. Consequently, the comprehension advantage for text/braille relative to audio domains may be due to the greater ease of making regressions in the former compared to the latter; indeed, previous research reveals that comprehension is supported by regressions that are made while reading (Schotter et al., 2014). Thus, if more frequent regressions are made with text and braille compared to auditory formats, comprehension should be significantly better and take significantly longer on average with text/braille vs. audio format.

While the studies presented in this dissertation could not explore this question directly, the time duration data for each presentation mode condition provide indirect information about the number of regressions made with each format. That is, there exists a positive relationship between the number of regressions one makes during reading and the amount of time it takes one to finish reading.

Interestingly, the results presented in Chapter 4 indicate that the average amount of time it takes sighted individuals to read a text is not significantly different than listening to a voice actor read the information aloud. This finding implies that a similar amount of regressions was made in each of these conditions. Additionally, the findings in Chapter 6 reveal that the average amount of time it takes blind individuals to read using a braille display is significantly longer compared to the hard-copy braille and audio conditions, and the hard-copy braille time duration is significantly longer than the two audio conditions. According to these results, blind individuals may have made more frequent regressions in both braille conditions compared to the auditory conditions. One should keep in mind, however, that the act of reading braille is slower than listening, so one must control for that in order to deduce the amount of time spent making regressions.

7.2 The trade-off between time duration and comprehension performance

The trade-off between time duration and comprehension ability was explored in Chapters 4, 5, and 6 of this dissertation by assessing the impact of presentation mode on the total assessment score divided by the amount of time taken in each format condition. Whereas there did not exist such a trade-off among sighted participants, blind participants' comprehension ability was significantly better in the auditory conditions after controlling for time duration.

These findings have direct implications for recommending presentation formats based on the goals and motivations of the reader. That is, if a blind individual wants to complete a reading task quickly, they should access the materials using an auditory mode rather than a braille format. If time is not a factor, however, then blind individuals should consider using a braille format for reading. For sighted individuals, text should generate better comprehension compared to audio regardless of the amount of time given to read. Taken together, these findings are presented in a way that is useful for individuals who are concerned with the amount of time spent reading materials.

7.3 The interaction between presentation mode and passage topic

As mentioned in Chapter 6, there was a significant interaction between the reading format and passage content; in other words, the effect of presentation mode on comprehension depends on the topic covered in the passage. Although sufficient data did not exist to statistically investigate this interaction, the passage topics and their trends are provided in Table 7.1 below.

Passage	Topic	Trend
Passage A	Zones of the Ocean	Better in both braille conditions versus audio conditions
Passage B	Relative dating of fossils; rock layers	Better in hard-copy braille versus braille display and audio conditions
Passage C	Viruses and vaccines	Ceiling effect across all formats; no clear differences
Passage D	Invertebrate symmetry	Better in braille display versus hard-copy braille and audio conditions

Table 7.1: Trends in impact of presentation mode on comprehension ability for each passage.

While the expectation was that both braille conditions would produce superior comprehension compared to both audio conditions for all passages, the trends presented in Chapter 6 and in Table 7.1 reveal otherwise. Interestingly, different braille formats generated better comprehension ability for Passage B and Passage D. A possibility is that the particular spatial movements required to use each of the braille formats may foster the creation of spatial representations in the same direction. Indeed, research has shown that improvements in spatial ability performance among sighted individuals occurs after training (Baenninger & Newcombe, 1989). Additionally, among blind individuals, spatial knowledge of the surrounding environment is enhanced after exploring a tactile map of it (Cattaneo et al., 2008; Espinosa et al., 1998). This effect demonstrates that blind individuals are able to utilize their experience with exploring a 2-dimensional spatial layout to improve their navigation of a 3-dimensional setting (Cattaneo et al., 2008; Espinosa et al., 1998).

As described in the first chapter of this dissertation, when using a refreshable braille display to read materials, individuals move their fingers laterally (left to right) across the braille until they reach the end of the line. Then, they press a button on the display that moves to the next physical line of the passage, as only one line is displayed at a time. Such constant horizontal movements may foster the reader's ability to create spatial representations along the horizontal direction; if so, this could possibly explain the superior performance in the braille display vs. other presentation modes for Passage D. To elaborate, Passage D considered radial and bilateral symmetry of invertebrates. Perhaps the focus on right-left mirror symmetry biased the results from a potential "training effect" produced by using the refreshable braille display.

While braille displays require predominantly lateral hand movements, reading hard-copy braille involves both horizontal movements across braille lines and vertical movements to read either subsequent or previous lines. Following logic similar to what is provided above, reading hard-copy braille may have primed the formation of vertical spatial representations, thereby bolstering understanding of topics that involve vertically-situated layouts. Such a topic is similar to Passage B, which discussed relative dating of fossils and rock layers. The rock layers discussed in the passage are vertical, where the rocks and fossils closest to the Earth's surface are considered younger compared to the ones closest to the core (Miller & Levine, 2008). Consequently, the vertical hand movements used to read the passage with hard-copy braille may have facilitated comprehension of Passage B, leading to the observed trend.

Passage A, which discusses zones of the ocean, does not place demands on spatial representations. This may explain the similar comprehension performance between the two braille conditions. Perhaps passages that do not require forming representations of a particular spatial layout are comprehended better with more actively-engaging formats (i.e., braille) compared to passive modes (i.e., audio). Furthermore, as discussed in Chapter 6, the ceiling effect discovered with Passage C could potentially be explained by greater familiarity with the topic, as viruses and vaccines are frequently discussed in the media, thereby exposing individuals to more information about this topic compared to the other passages used for the studies presented here.

As the considerations presented in this section are merely speculation, future research should directly investigate them in order to clarify the relationship between passage topics and reading formats and how this relationship effects comprehension ability.

7.4 Educational implications

7.4.1 Experimental design and methods enhance educational implications

The experimental methodology used in the studies presented here strengthened the ability to directly apply the results to inform education. A novel comprehension assessment was developed to directly compare the results between blind and sighted individuals. Not only did making such a comparison aim to elucidate the underlying comprehension mechanism, but also it emphasized the need to develop assessment tools that are appropriate for groups of individuals that may have different cognitive abilities. Using the same assessment for groups that potentially differ in this way could lead to misleading results and implications, which may be deleterious to academic growth.

These studies also incorporated currently-used assistive technology. As these devices are oftentimes more convenient and transportable, investigating the trade-off between comprehension and accessibility is essential. Additionally, these studies allowed for personalized adjustments to speech settings and for making regressions in both audio and braille/text. Previous research on this topic has not allowed for regressions during audio conditions, even though individuals made regressions during text/braille conditions. Since participants were able to use their preferred settings and to make regressions, this dissertation studies represented a setting more reminiscent of the real-world, thereby enhancing the applicability of the results to education.

Furthermore, these studies used a within-subjects design, which produced more substantive results than a between-subjects design, because the comprehension differences between presentation modes were not biased due to varying preferences and/or experience with the different modes and devices. By using a within-subjects design, each participant acted as their own "control," as preference and experience level were held constant across presentation modes.

7.4.2 Meaningful comprehension differences between presentation modes

As shown in Chapters 4-6, the comprehension differences between presentation modes that are significant may not seem very large in magnitude. Take, for instance, the significant comprehension difference between text and audio formats among sighted participants presented in Chapter 4. Although the difference is significant, reading materials in text format only produces a 5% comprehension advantage compared to listening to them in audio format. This difference magnitude may seem small; however, considering it in the context of an educational setting elucidates its importance.

To illustrate, an individual studying for a biology exam wants to receive a passing grade. The comprehension results presented in Chapter 4 reveal that a 5% difference in scores achieved between text and voice actor formats is meaningful insofar as reading yields an average grade of "C" (i.e., 74.48%), whereas listening yields an average grade of "D"(i.e., 69.7%). Indeed, the amount of meaning behind this interpretation assumes a particular motivation: that the individual aims to receive a passing grade. As this is a typical motivating factor in educational settings, the 5% difference in comprehension scores is considered meaningful.

7.4.3 Considering other assistive technologies for accessing reading materials

While the findings presented here highlight the benefit of using braille formats to access reading materials, they do not recommend discounting the use of audio-based assistive technology to access reading materials. The present results involving blind individuals particularly apply to fluent braille readers' comprehension of high-school biology passages using four different presentation modes. They do not suggest that reading comprehension is impaired by utilizing audio-based assistive technology in general. Indeed, it may be challenging for individuals who become blind later in life to learn braille, so they depend heavily on auditory means of accessing reading materials. There exist many other audio-based assistive technologies not tested here that could produce different results with the present experimental methodology. That said, the findings presented in this dissertation are not meant to serve as a strict guideline for informing individuals how to access their reading materials, as this is ultimately at the discretion of the individual; yet, the results suggest that braille should not be replaced, but rather supplemented, by audio-based assistive technology in educational settings.

7.5 Challenges of creating assessment tools

Many challenges come with developing assessment tools that accurately assess comprehension. To control for passage difficulty, the Flesch-Kincaid Readability Test (Kincaid et al., 1975) was performed on each passage prior to experimentation. Nevertheless, there still existed significant differences in comprehension between passage topics for both blind and sighted individuals.

One way to prevent these effects a-priori is by collecting sufficient pilot data; however, at least 30 of each blind and sighted participants were needed to generate reliable fit statistics for the comprehension questions via Rasch modeling (see Chapter 2 for additional information). Unfortunately, it was exceedingly difficult to recruit fluent braille readers, thereby limiting the ability to test enough participants for pilot data. Despite this shortcoming, this study was the first of its kind to take appropriate statistical measures to analyze questions that accurately measured comprehension for each group. Interestingly, the fit statistics for blind and sighted individuals were different, further emphasizing the need to consider differences in participant groups when developing assessments.

7.6 Future directions

Although these studies aimed to reveal the underlying mechanism of comprehension, they were unable to provide a clear answer and highlighted several empirical questions that remain to be addressed. First, future research on this topic should investigate the impact of presentation mode on regression frequency. Although regressions were not directly measured in the studies presented here, the longer time duration found in braille compared to audio conditions may, in part, be due to more frequent regressions (see Chapters 5 and 6). If more frequent regressions do occur when reading braille compared to listening to audio, future research should investigate the reason underlying this difference. One possible explanation could be that the Euclidean geometry and the spatial layout of braille facilitates regressions compared to auditory devices. If this is the case, then research should consider ways to promote regressions for auditory formats, as regressions are associated with improving comprehension (Schotter et al., 2014).

Future studies should consider using similar experimental methodology presented here to test additional assistive technologies. It would also be insightful to explore different types of synthesized speech and investigate their impact on understanding. As mentioned in Chapter 3, the studies here employed NVDA with an Eloquence add-on to improve voice quality; however, it was still critiqued by blind users as being unclear when compared to newer software that cost money, such as JAWS (Freedom Scientific, 2009).

Additionally, while the present experiments used non-fiction, biology passages, future studies should investigate the effect of presentation mode on comprehension for literary passages. A potentially interesting consideration would be to alter the amount of dialogue presented in each passage to explore whether this manipulation impacts the results. That is, perhaps passages with more dialogue are better understood with auditory formats, and as the presence of dialogue is decreased, the comprehension benefit provided by audio is mitigated. In addition to testing different types of passages (e.g., literary), studies should also examine the relationship between the type and amount of spatial content in a passage and the hand motion required to read, as described in the section above regarding the interaction between presentation mode and passage topic.

Finally, it would be worthwhile to extend these studies to include individuals with low vision. Since many of these individuals have residual vision, they utilize different assistive technologies to access reading materials, such electronic video magnifiers, closed circuit television systems (CCTVs), typoscopes, and the OrCam device (Gerritsen & Duffy, 2018; Na'aman et al., 2017). Results from such studies will have implications for a more diverse population while possibly providing further insights into the neural underpinnings of comprehension ability.

Bibliography

Adams, R. J. & Khoo, S. T. (1993). Quest : The interactive test analysis system. *Australian Council for Educational Research*.

Allen, M. C. (2008). Neurodevelopmental outcomes of preterm infants. *Current Opinion in Neurology*, *21*(2), 123–128.

American Foundation for the Blind (n.d.). Braille printers. *https://bit.ly/2JBkIGD*.

Armstrong, E. (2003). Braille version of 'harry potter' weighs in at 13 volumes. *https://www.csmonitor.com/2003/0701/p12s01-lecl.html*.

Arneson, A., Freund, R., Toyama, Y., & Wilson, M. (2015). Construct mapping with rasch, crasch package for r. *GitHub Repository; github.com/amyarneson/crasch*.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*.

Baenninger, M. & Newcombe, N. (1989). The role of experience in spatial test performance: a meta-analysis. *Sex roles*, *20*, 327–344.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using eigen and s4. *R package version 0.999999-2*.

Bell, E. C. & Mino, N. M. (2015). Employment outcomes for blind and visually impaired adults. *Journal of blindness innovation and research*, *5*(2).

Beni, R. D. & Cornoldi, C. (1985). Effects of the mnemotechnique of loci in the memorization of concrete words. *Acta Psychologica*, *60*.

Birch, E. E. & O'Connor, A. R. (2001). Preterm birth and visual development. *Seminars in Neonatology*, *6*(6), 487–497.

Birenbaum, M. & Tatsuoka, K. K. (1987). Open-ended versus multiple-choice response formats—it does make a difference for diagnostic purposes. *Applied Psychological Measurement*, *11*(4), 385–395.

Brysbaert, M. & Stevens, M. (2018). Power analysis and effect size in mixed effects models: a tutorial. *Journal of cognition*, *1*(1).

Carretti, B., Borella, E., & Beni, R. D. (2007). Does strategic memory training improve the working memory performance of younger and older adults? *Experimental psychology*, *54*, 311–320.

Cattaneo, Z., Fatino, M., Silvanto, J., Tinti, C., Pascual-Leone, A., & Vecchi, T. (2010). Symmetry perception in the blind. *Acta psychologica*, *134*, 398–402.

Cattaneo, Z., Vecchi, T., Cornoldi, C., Mammarella, I., Bonino, D., Ricciardi, E., & Pietrini, P. (2008). Imagery and spatial processes in blindness and visual impairment. *Neuroscience and behavioral reviews*, *32*, 1346–1360.

Chang, C. K., Wange, J. D., Tang, H. P., Cheng, C. M., & Lin, C. Y. (2014). Psychometric evaluation using rasch analysis of the whoqol-bref in heroin-dependent people undergoing methadone maintenance treatment: further item validation. *Health Qual Life Outcomes*.

Chien, T., linacre, J. M., & Wang, W. (2011). Examining student ability using kidmap fit statistics of rasch analysis in excel. *Advances in Information Technology and Education, International Conference, CSE 2011, Proceedings Part I*.

Clark, A. (1999). An embodied cognitive science? *Trends in cognitive sciences*, *3*, 345–351.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York: NY: Routledge Academic.

Craik, F. I. & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of experimental psychology*, *104*, 268.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 397–334.

Dandona, R. & Dandona, L. (2001). Refractive error blindness. *Bulletin of the World Health Organization*, *79*, 237–243.

D'Andrea, F. M. (2009). From carvings to computers: a history of tactile codes for people who are blind. In *The Braille Code Past Present and Future*. The International Council for Education of People with Visual Impairment.

Dehaene, S. & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in cognitive sciences*, *15*, 254–262.

Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2011). Fortune favors the (): Effects of disfluency on educational outcomes. *Cognition*, *118*, 111–115.

Dietrich, G. (n.d.). Presentation: Daisy 101. *https://bit.ly/2CBjZzc*.

Dixon, J. (2000). *Braille into the next millennium.* Washington D.C.: National Library Service for the Blind and Physically Handicapped and Friend of Libraries for Blind and Physically Handicapped in North America.

Dixon, R. A., Simon, E. W., Nowak, C. A., & Hultsch, D. F. (1982). Text recall in adulthood as a function of level of information, input modality, and delay interval. *Journal of gerontology*, *37*(3), 358–364.

Espinosa, M. A., Ungar, S., Ochaita, E., Blades, M., & Spencer, C. (1998). Comparing methods for introducing blind and visually impaired people to unfamiliar urban environments. *Journal of environmental psychology*, *18*, 227–287.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using g*power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149–1160.

Faul, F., Erdfelder, E., Lang, A.-G., , & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*.

Fleiss, J. L., Cohen, J., & Everitt, B. S. (1969). Large sample standard errors of kappa and weighted kappa. *Psychological Bulletin*, *72*(5), 323.

Forster, S. & Lavie, N. (2009). Harnessing the wandering mind: The role of perceptual load. *Cognition*, *111*, 345–355.

Foulke, E., Amster, C. H., Nolan, C. Y., & Bixler, R. H. (1962). The comprehension of rapid speech by the blind. *Exceptional Children*, *29*(3).

Freedom Scientific (2009). Jaws screen reader. *http://www.freedomscientific.com/products /fs/jawsproduct-page.asp.*

Gerritsen, B. & Duffy, M. A. (2018). Overview of low vision devices. *https://bit.ly/2CBlLk1.*

Good, W. V., Jan, J. E., Burden, S. K., Skoczenski, A., & Candy, R. T. (2001). Recent advances in cortical visual impairment. *Developmental Medicine and Child Neurology*, *43*, 56–60.

Green, R. G. (1981). Remembering ideas from text: the effect of modality on presentation. *British Journal of Educational Psychology*, *51*, 83–89.

Hackett, A. (2016). A low-cost revolution in refreshable braille. *Perkins School for the Blind*, http://www.perkins.org/stories/a–low–cost–revoluation–in–refreshable–braille.

Hinton, P. R., McMurray, I., & Brownlow, C. (2004). *SPSS explained.* Routledge.

Irwin, R. B. (2009). The war of the dots. *American Foundation for the Blind.*

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive bu largely ignored problem. *Journal of personality and social psychology, 103*(1).

Kincaid, J. P., Jr., R. P. F., Rogers, R. L., & Chissom, B. S. (1975). Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel (no. rbr-8-75). *Naval Technical Training Command Millington TN Research Branch.*

Kingman, S. (2004). Glaucoma is second leading cause of blindness globally. *Bulletin of the World Health Organization, 82,* 887–888.

Kingstone, A., Smilek, D., & Eastwood, J. D. (2008). Cognitive ethology: a new approach for studying human cognition. *British Journal of Psychology, 99,* 317–340.

Kintsch, W., Mandel, T. S., & Kozminsky, E. (1977). Summarizing scrambled stories. *Memory and cognition, 5*(5), 547–552.

Legge, G. E., Madison, C., Vaughn, B. N., Cheong, A. M., & Miller, J. C. (2008). Retention of high tactile acuity throughout the life span in blindness. *Perception and psychophysics, 70,* 1471–1488.

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). Package 'lsmeans'. *The American Statistician, 34*(4).

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Lowenfeld, B. (1945). Braille and talking book reading: a comparative study. *Dissertation.*

Lowenfeld, B., Abel, G. L., & Hatlen, P. H. (1968). *Blind children learn to read.* 301-327 East Lawrence Avenue, Springfield, Illinois: Charles C. Thomas.

Millar, S. (1997). *Reading by touch.* 11 New Fetter Lane, London and 29 W. 35th Street, New York: Routledge.

Miller, K. R. & Levine, J. S. (2008). *Miller Levine Biology Student Edition.* Pearson Education, Inc. or its affiliates. Used by permission. All rights reserved.

Na'aman, E., Shashua, A., & Wexler, Y. (2017). User wearable visual assistance system.

National Federation for the Blind (2007). A parent's guide to the slate and stylus. *https://nfb.org/images/nfb/publications/fr/fr25/fr07spr18.htm.*

National Federation for the Blind (2012). Legally blind definition. *https://bit.ly/2uq7zpl.*

National Library Service for the Blind and Physically Handicapped at the Library of Congress (n.d.). Nls at the library of congress homepage. *https://www.loc.gov/nls.*

National Library Service for the Blind and Physically Handicapped, Library of Congress (2008). Specification 800. *Braille books and pamphlets.*

Nolan, C. Y. (1963). Reading and listening in learning by the blind. *Exceptional Children, 29*, 313–316.

Nunnally, J. (1978). Psychometric methods. *New York: McGraw-Hill.*

Pearson (2009). *Wechsler Individual Achievement Test-Third Edition.* San Antonio, TX: Pearson.

R Core Team (2018). *R: A Language and Environment for Statistical Computing: Version 3.4.4.* Vienna, Austria: R Foundation for Statistical Computing.

Roder, B., Rosler, F., & Neville, H. J. (2001). Auditory memory in congenitally blind adults: a behavioral-electrophysiological investigation. *Cognitive Brain Research, 11*, 289–303.

Rogowsky, B. A., Calhoun, B. M., & Tallal, P. (2016). Does modality matter? the effect of reading, listening, and dual modality on comprehension. *SAGE Open.*

Roman, C., Baker-Nobles, L., Dutton, G. N., Luiselli, T. E., Flener, B. S., Jan, J. E., Lantzy, A., Matsuba, C., Mayer, D. L., Newcomb, S., & Nielsen, A. S. (2008). Statement on cortical visual impairment. *American Foundation for the Blind.*

Russomanno, A., O'Modhrain, S., Gillespie, B. R., & Rodger, M. W. (2015). Refreshing refreshable braille displays. *IEEE Transactions on Haptics, 8*(3), 287–297.

Schotter, E. R., Tran, R., & Rayner, K. (2014). Don't believe what you read (only once) comprehension is supported by regressions during reading. *Psychological science, 25*(6), 1218–1226.

Smallwood, J., Fishman, D. J., & Schooler, J. (2007). Counting the cost of an absent mind: mind wandering as an underrecognized influence on educational performance. *Psychonomic Bulletin and Review, 14*, 230–236.

Supalo, C. A., Isaacson, M. D., & Lombardi, M. V. (2013). Making hands-on science learning accessible for students who are blind or have low vision. *Journal of Chemical Education, 91*(2).

Thomson, D. R., Besner, D., & Smilek, D. (2013). In pursuit of off-task thought: mind wandering-performance trade-offs while reading aloud and color naming. *Frontiers in Psychology, 4*, 1–11.

Varao Sousa , T. L., Carriere, J. S. A., & Smilek, D. (2013). The way we encounter reading material influences how frequently we mind wander. *Frontiers in psychology, 4*(892), 1–8.

Vohr, B. R., Wright, L. L., Dusick, A. M., Lisa Mele, J. V., Steichen, J. J., Simon, N. P., Wilson, D. C., Broyles, S., Bauer, C. R., Delaney-Black, V., Yolton, K. A., Fleisher, B. E., Papile, L.-A., & Kaplan, M. D. (2000). Neurodevelopmental and functional outcomes of extremely low birth weight infants in the national institute of child health and human development neonatal research network. *Pediatrics*, *105*(6), 1216–1226.

Wechsler, D. (2008). *Wechsler adult intelligence scale - fourth edition*. San Antonio, TX: The Psychological Corporation.

Wiederholt, J. L. & Bryant, B. R. (2012). *Gray Oral Reading Test-Fifth Edition (GORT-5): Examiner's Manual*. Austin, TX: Pro-Ed.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic bulletin and review*, *9*, 625–636.

Wilson, M. (2005). *Constructing measures: an item response modeling approach [E-reader version]*. https://ebookcentral-proquest-com.libproxy.berkeley.edu/lib/berkeley-ebooks.

Winter, B. (2013). Linear models and linear mixed effects models in r with linguistic applications. *arXiv preprint arXiv:1308.5499*.

Woolley, G. (2011). Reading comprehension. In *Reading comprehension: Assisting children with learning difficulties*. Springer Netherlands.

World Health Organization (2016). Visual impairment and blindness. *https://bit.ly/2HFcs6J*.

Yerkes, R. M. & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of comparative neurology and psychology*, *18*, 459–482.

Appendix A

Passages and questions

A.1 Practice passages and questions

As mentioned in Chapter 3 of this dissertation, the practice passages were obtaind from the Gray's Oral Reading Fluency Test, Fifth Edition (Wiederholt & Bryant, 2012). They written at a fourth to fifth grade reading level, according to the Flesch-Kincaid Readability Test (Kincaid et al., 1975)

A.1.1 Voice actor practice

The following passage was used to adjust the voice rate and volume to the preferred settings. Participants were also instructed to press the left and right keys on the keyboard to rewind and fast forward the audio. No questions were asked after this passage.

Passage: "Twelve girls were playing a game at a party. On the wall before them hung a picture of a lion with a fierce look in his eye. The girls first put large paper bags over their heads so they couldn't see. As the clock ticked off one minute after another, each of the girls tried to pin a ribbon on the lion's tail. They put ribbons on the lion's paws, head, and body, which gave the lion a friendly look. Everyone missed the tail, so nobody won the prize. After the game was finished, everyone had cake and ice cream."

This next practice passage was played at the voice rate the participant selected during the practice listed above. The participants were instructed to listen to this passage for comprehension, using the left and right keys if desired, and that it would be followed by two free-response comprehension questions.

Passage: "It was pet day at the fair. Under the big pink tent the children were waiting for the parade of animals to begin. They had trained their pets to do many different tricks in front of the judges. One girl had a bunny that could climb a ladder, and another had a talking bird. Among the youngsters was a tall, red-headed boy whose goat made trouble for him. He had trained it to dance, but it kicked and tried hard to break away. When it heard the band though, it became still. During the parade, it danced so well it won a prize—a blue ribbon for first place."

Questions:

1. What color was the tent where the children were waiting?
2. According to the passage, what trick could the bunny perform?

A.1.2 Text or braille practice

The following passage was used to familiarize the participants with the text or braille condition. Participants were instructed to read the following passage for comprehension and that it would be followed by two free-response comprehension questions.

Passage: "Once there was a turtle who chattered so much that she had no friends. One day she met an eagle traveling to faraway lands across the sea. The turtle had always wished for an adventure. It would make a wonderful tale and earn her many friends. So the eagle agreed to take the turtle along. He flew with a strong stick in his claws while the turtle held the stick fast in her mouth. But when they were far out to sea, the turtle could no longer keep silent. As soon as she opened her mouth, she fell straight into the ocean below."

Questions:

1. Why did the turtle want to go with the eagle?
2. Why doesn't the turtle have any friends?

A.1.3 Screen reading practice

The following passage was used to adjust the speech rate and volume to the preferred settings. Participants were also introduced to the different key combinations they could use to navigate through the passage. No questions were asked after this passage, as this practice session was meant to focus on preference and technology training.

Passage: "A greedy fox was eating a large bowl of soup when a stork came to call. Disappointed at having to share his supper, the fox gave the stork a bit of the soup in a shallow dish. But the stork could not eat from the dish. Some days afterward, the stork invited the fox to dinner. She had made some delicious soup, and the fox was very hungry. But she served the soup in two tall bottles with thin, narrow necks. The stork stuck her long bill into the bottle and ate every drop, while the fox could do nothing but watch."

This next passage was played at the speech rate the participant selected during the practice listed above. The participants were instructed to listen to this passage for comprehension, using the key combinations they learned, and that it would be followed by two free-response comprehension questions.

Passage: "A blue jay was perched on a limb looking for water. Having just flown a great distance, she was very thirsty. At that moment she happened to spot a water jar on the ground, so she flew down and tried to get a drink from the jar. But there was so little water in the jar that she was unable to drink. Just as she felt that she would surely die of thirst, an idea struck her. The jay gathered a pile of stones and began dropping them in the jar. Little by little the water rose, and at last the jay could drink her fill."

Questions:

1. What is the primary challenge in this story?
2. Why couldn't the jay drink the water?

A.1.4 Refreshable braille display practice

The following passage was used to introduce the participants to the layout of the braille display and how to use the keys to advance and regress through the passage. No questions were asked after this passage.

Passage: "An elephant was walking through the jungle alone, frightened that he might come upon a snake. He had never seen a snake before, but he had heard what terrible creatures they were. He had not gone very far when a snake with green and yellow stripes crossed his path. Instantly the elephant froze in his tracks. At that moment a hunter stepped out of the bushes, his gun pointing straight at the elephant. But when the hunter caught sight of the snake, he dropped his gun and fled. That was how the elephant learned that snakes were not what he had thought."

Participants were instructed to use the braille display to read the next passage for comprehension, as they would be asked to answer two free-response questions after reading it.

Passage: "One bright summer day, a young boy and his grandmother walked to a nearby pond to fish. The boy's grandmother showed him how to put worms on the hook so they would not come off. For a long while, they sat quietly waiting for the fish to bite. Suddenly the boy got a good bite. As he tried to land the fish, he became so excited that he dropped his pole into the water. The fish quickly swam away with it, and soon the pole had disappeared. The boy looked wide-eyed at his grandmother. Then they both had a good laugh."

Questions:

1. What did the grandmother show the boy how to do?
2. What one word did the story use to describe how the boy felt when he tried to land the fish?

A.2 Experimental passages and questions

Permission from Pearson Publishing was acquired to utilize the experimental passages listed below; they were acquired from the Miller Levine Biology Student Edition 2008C by Kenneth R. Miller and Joseph S. Levine (Miller & Levine, 2008, Copyright Pearson Education, Inc. or its affiliates). The experimental passages were written at a 12th grade reading level (Kincaid et al., 1975).

As mentioned in Chapter 2, originally, each participant answered a total of eight free-response questions, and the order of these questions was randomized across participants. The infit statistics generated from the statistical analyses described in Chapter 2 led to the elimination of two questions per experimental passage for the final data analysis (i.e., the analysis that assessed comprehension differences between presentation modes). The questions that were removed from the final data analyses according to their infit statistics are presented in parentheses below. Additionally, recall that each passage was presented in each presentation mode condition, as the presentation mode-experimental passage pairings were counterbalanced and randomized across participants.

Passage A: "Unless you are an avid diver or snorkeler, it takes some imagination to picture what life is like in the vast, three-dimensional ocean. Sunlight penetrates only a relatively short distance through the surface of the water. Photosynthesis is limited to this well-lit upper layer known as the photic zone. Only in this relatively thin surface layer—typically down to a depth of about 200 meters—can algae and other producers grow. Below the photic zone is the aphotic zone, which is permanently dark. Chemosynthetic autotrophs are the only producers that can survive in the aphotic zone. There are several different classification systems that scientists use to describe marine ecosystems. In addition to the division between the photic and aphotic zones, marine biologists divide the ocean into zones based on the depth and distance from shore."

Passage A Questions:

1. What is the difference between the photic and aphotic zones of the ocean?
2. According to the passage, in addition to the division between photic and aphotic zones, what are some other factors that marine biologists use to divide the ocean into zones?
3. (Why is the photic zone so shallow?)
4. Approximately how deep is the photic zone?
5. What do you think is necessary for photosynthesis, given where it occurs in the ocean based on the passage?
6. According to the passage, chemosynthetic autotrophs are the only producers that live in the aphotic zone. Why do you think this is?
7. (What do you predict would happen to algae in the aphotic zone? Why?)

8. An avid diver is exploring the ocean and needs a flashlight to observe her surroundings. How deep do you think the diver is?

Passage B: "About two centuries ago, geologists noted that rock layers containing certain fossils consistently appeared in the same vertical order no matter where they were found. Also, a particular species of trilobite—a common fossil and an extinct relative of horseshoe crabs—might be found in one rock layer but be absent from layers above or below it. How might such a pattern be useful? In relative dating, the age of a fossil is determined by comparing its placement with that of fossils in other layers of rock. Sedimentary rock is formed from the gradual deposition of layers of sand, rock, and other types of sediment. The rock layers form in order by age—the oldest layers on the bottom, with more recent layers on top, closer to Earth's surface."

Passage B Questions:

1. What is relative dating?
2. (According to the passage, what is trilobite?)
3. How are sedimentary rocks formed?
4. What did geologists note about rock layers containing certain fossils about two centuries ago?
5. How accurate do you think relative dating is in determining a fossil's age?
6. If an igneous rock layer cuts across metamorphic rock layers, what does this say about the age of the igneous rock relative to the metamorphic rock?
7. If Sedimentary Layer A occurs above Sedimentary Layer B, which layer is older?
8. (If a mammal's fossil is found in the 10th rock layer in Australia, which layer would you suggest a United States archaeologist should investigate if she is searching for the same mammal?)

Passage C: "Like bacteria, viruses produce disease by disrupting the body's normal equilibrium. In many viral infections, viruses attack and destroy certain cells in the body, causing the symptoms of the disease. Poliovirus infects and kills cells of the nervous system, producing paralysis. Other viruses cause infected cells to change their patterns of growth and development. Unlike bacterial diseases, viral diseases cannot be treated with antibiotics. The best way to protect against most viral diseases lies in prevention, often by the use of vaccines. Several decades of childhood vaccination against smallpox, a terrible viral disease that once killed millions, have virtually eliminated this disease. Most vaccines provide protection only if they are used before an infection begins. Once a viral disease has been contracted, it may be too late to control the infection. However, sometimes the symptoms of the infection can be treated."

Passage C Questions:

1. What is one way that viruses disrupt the body's normal equilibrium?
2. According to the passage, what is the best way to protect against most viral diseases?
3. Why does poliovirus lead to paralysis?
4. (Based on the passage, how do bacteria produce disease?)
5. Why do you think that most vaccines are given to young children and not to adults?
6. (A man with smallpox goes to the doctor and is informed that he should take an antibiotic. Do you predict that the man will recover? Why or why not?)
7. Why do you think that smallpox killed millions of people in the past?
8. Is it possible for someone to have no symptoms of a particular viral infection yet also have that virus in their system? Why?

Passage D: "All invertebrates except sponges exhibit some type of body symmetry. Jellyfish and starfish exhibit radial symmetry—body parts extend from the center of the body. Worms, mollusks, and arthropods exhibit bilateral symmetry, or have mirror-image left and right sides. Most invertebrates with bilateral symmetry rely on movement for feeding, defense, and other important functions. The evolution of this body plan and lifestyle was accompanied by the trend toward cephalization, which is the concentration of sense organs and nerve cells in the front of the body. Invertebrates with cephalization can respond to the environment in more sophisticated ways than can simpler invertebrates. In most worms and arthropods, nerve cells are arranged in structures called ganglia. In more complex invertebrates, such as certain mollusks, nerve cells form an organ called a brain."

Passage D Questions:

1. What is the difference between radial and bilateral symmetry?
2. What does cephalization mean?
3. Name two out of the three invertebrates that are listed in the passage as having bilateral symmetry.
4. (Where are nerve cells arranged in most worms and arthropods?)
5. Why do you think that sponges, a primitive sea-floor organism, do not exhibit cephalization?
6. Which form of symmetry, bilateral or radial, does the author imply leads to more sophisticated animals?

7. (Based on the passage, does the author believe that jellyfish and starfish are sophisticated creatures?)
8. Using the terms that you read about in the passage, how would you categorize a snail?