ABSTRACT

Title of Dissertation: TEXT COMPREHENSION ACROSS PRINT AND AUDIO: A PERSON-CENTERED MIXED METHODS STUDY

Dissertation Directed by: Dr. Patricia A. Alexander
Department of Human Development and Quantitative Methodology

The use of the audio medium (e.g., audiobooks and podcasts) is proliferating in everyday and educational contexts. Yet, research investigating text processing in audio compared to the more commonly used print medium is limited in scope. Specifically, the research so far has majorly focused on younger learners or English language learners, narrative genre texts, operationalized comprehension as a unidimensional construct, and used variable-centered analytical techniques.

The current mixed methods study aimed to explore text processing across print and audio by focusing on four interrelated dimensions—learner, text, task, and test. I used finite mixture modeling for the quantitative part of the study to identify meaningful reader and listener profiles. Following the identification and validation of profiles, students from the profiles were interviewed to complement and enhance the understanding of the groups. Specifically, the study aimed to investigate differences across the two mediums vis-à-vis learner characteristics, text and test processing behaviors, and comprehension outcomes. A further goal of the study was to identify meaningful and distinct reader and listener profiles by accounting for affective and behavioral variables, and validating the profiles on cognitive variables. Finally, the study aimed to build qualitatively rich descriptions of the quantitatively unearthed profile groups.

To address these aims, undergraduate students (n =130) were recruited from human development courses. They completed measures related to self-efficacy and reported their reading and listening habits. Each participant's screen was recorded as they processed text in print and audio. Text processing behaviors (e.g., scrolling, increasing playback speed) and off-task behaviors (e.g., eating, fidgeting) were coded. Learner-related, text processing, and task variables were used to find meaningful reader and listener profiles. The profiles were validated using prior topic knowledge and comprehension as covariate and outcome, respectively.

Students belonging to each profile were invited for interviews (n = 10). The format was a cued retrospective interview, wherein video clips were used to prompt participants. The interviews were transcribed, segmented into utterances, and coded for learner-related, text-related, task-related, and test-related content.

Results from the variable-centered analysis revealed that reading print or listening to audio led to similar performance levels on items targeting recall and inference. However, reading print was associated with higher scores on the item assessing the main idea than listening to audio. Results from the mixture modeling and interviews revealed three reader profiles—*Distracted Surfers, Labored Harvesters, and Fluent Surveyors*—and three listener profiles—*Inconsistently Attentive, Inattentive, and Persistently Attentive*. The profiles were found to differ qualitatively on strategies, text processing depth, and attention regulation.

This study's contribution is in expanding the research on comprehension across different mediums both in terms of scope and methodologically. The current investigation demonstrates that learner characteristics and text processing behaviors need to be

accounted for when studying comprehension with different mediums. Practically, it has implications for practitioners looking to incorporate audio for content delivery in their courses and for instructional designers developing educational technology tools to optimize learning.

TEXT COMPREHENSION ACROSS PRINT AND AUDIO: A PERSON-CENTERED MIXED METHODS STUDY

by

Anisha Singh

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2023

Advisory Committee:

Dr. Patricia A. Alexander, Chair
Dr. Doug Lombardi
Dr. Min Wang
Dr. Gregory R. Hancock
Dr. Sarah McGrew
Dr. Yasmeen Faroqi-Shah (Dean's Representative)

Acknowledgments

I have dreamed this dream in all its beautiful versions. But now, as I sit down to thank all those who made this Ph.D. a reality, I fear that I will not do justice in conveying the depth of my gratitude. So, I begin with an apology for not having the words to tell you all how much your support and love have meant to me.

Patricia, how do I even begin? You have imparted an enduring passion for the pursuit of science. You have shown me how to weed out the unnecessary and "see" what is right in front of us. You have taught me that, more often than not, the questions worth asking are the simplest ones. And most importantly, you have modeled that love and playfulness are the necessary ingredients for living a life of passion and purpose.

Lab friends, I would not be here without our lively discussions, your piercing questions, and your supportive suggestions. You pushed me as a researcher to flesh out my vague ideas and were the most helpful and kind critics. This dissertation study would not have seen the light of day were it not for all of you. A special thanks to Jules, Eric, and Jannah for facilitating the data collection. And to dearest Yuting, for making yourself available at all hours when I needed a friend to talk to about research or my emotions. To my friend, Sophie, you are the reason I applied to this program, and your insights and perspectives always brought a welcome freshness. And finally, my roommates, who have been more like family—Suzy, Daniel, Selu, and Stefano—thank you for making this journey so enjoyable.

My research assistant, Blaine Hoffman, for bringing your curiosity to this project. Thank you for painstakingly coding all the videos and establishing interrater reliability with me.

The brilliant faculty, you have given me multiple lenses and tools to dissect the world. You also pushed the boundaries of my aspirations by sharing your expertise and showing me that there are many ways of asking and answering one question. Dr. Min Wang, thank you for imparting your methodic approach to science. As I launch my research career, I know the foundation you helped me build will be critical to my success. Dr. Greg Hancock, you unlocked my statistical imagination by helping me visualize abstract concepts (yes, I am referring to the sticks!). You have given me new ways of thinking. Dr. Doug Lombardi, thank you for always keeping your (virtual) door open to discuss research and gently correcting me when I was going astray. Dr. Sarah McGrew, I have leaned on your research to guide me in my qualitative endeavors. Your scholarship and its impact have been an inspiration. Dr. Yasmeen Faroqi-Shah, you have known me since when I started the program, and I have enjoyed working with you. I am grateful for your kindness.

My husband, Zach, thank you for giving me strength, endless hugs, believing in me, showing me that reason defeats fear, and for feeding me! Your unwavering emotional and intellectual support has helped me achieve my dream. Thank you for being my companion. My baby, Noor, your laughter, pure joy, and how you look at me overshadows the value of everything but makes me want to achieve everything to be my best version for you.

My parents, you are my constants, I know that there is no one who sees me the way you do, and I am grateful that you gave me the independence to find my way. My sister, Sam, you know a part of me that no one does, you have assuaged my fears since I was a little girl and given me the confidence to achieve my dreams. My darling nephew, Ivaan, your joyful energy has been a salve whenever I have felt down. To my Esmili family, your

contribution to this journey has been in the form of your deep love for ideas and your healthy skepticism. Finally, to my grandmother, thank you for imparting in all your grandchildren a spirit of independence and fearlessness.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I: INTRODUCTION

The audio medium is experiencing a renaissance (Pew Research Center, 2021, 2022). Downloading and streaming audiobooks and podcasts with mobile digital devices has allowed unlimited access to texts on audio (Spotify, 2022). The Pew Research Center attests to this spurt in audio usage. They reported that one in five Americans had listened to at least one audiobook in the year surveyed. Within the population of college graduates, the proportion of individuals using audiobooks was even higher at 34% (Pew Research Center, 2019). The popularity of audio is reflected in its sales as well. According to Audio Publishers Association, there has been double-digit growth in sales year on year in the last decade (Audio Publishers Association, 2021). During the same time, the sales of e-books and paper books have been declining (Pew Research Center, 2019).

There has been a parallel surge in the use of the audio medium in academic contexts. This medium is no longer viewed only as a support for students with dyslexia or visual impairments (Noland, 2020). Nowadays, many advanced academic readings are available in audio format. For example, digital textbooks for college students and online journal articles increasingly contain a "read aloud" feature (e.g., Taylor and Francis Online, vitasource.com). Using audio as the delivery medium of textual content has also been described as a valuable instructional resource for students in developing countries who might have limited access to literacy opportunities (Friedland et al., 2017; School-to-School International, 2017).

The history of audio as a conduit for information is long. By the 1920s, radio made its debut as a popular broadcast medium. From the start, radio included

dramatizations of stories, viewed as a precursor to the modern-day podcast (Baron, 2021). Like many other inventions in human history, audiobooks are also linked to war. Audiobooks came into being the years following World War I as sound recordings of books for soldiers who had been blinded during the War. Fast forward to the 1970s, and Books on Tape, a company that rented out cassettes of full-length books, was born. Mainly targeted toward commuters in cars at the time, Books on Tape, experienced another surge in popularity with the advent of the Sony Walkman®. In the 1980s, audiobooks could now be listened to while on foot (Rubery, 2016). Between the 1990s and early 2000s, the use of audio as a medium for text experienced a brief decline before making a comeback in a new digital form.

Research Landscape

The research on the viability of audio as a medium for text follows a similar historical trajectory. One of the early reviews of the literature conducted by Sticht and colleagues (1974) covered research from 1917 to 1968. The purpose of the review was to chart the developmental relation between reading and listening comprehension. The authors concluded that younger students (until grades 6–7) comprehend more with audio than printed text. Beyond seventh grade, they reported an advantage for reading over listening when it comes to comprehension. Arguably, the affordances of the digital audio medium, its usage context, and the type and quality of content available now are markedly different from the first half of the 20th century. This digital revolution necessitates a renewed investigation into the audio medium and how it compares to its print counterpart.

Two recent reviews shed light on the state of research comparing audio and print. Clinton (2021) meta-analyzed 46 studies conducted between 1955 to 2020. The primary purpose of the meta-analysis was to compare reading and listening comprehension. She found no statistically significant difference between reading and listening comprehension. Findings from the moderator analyses showed that reading led to better outcomes on inferential comprehension questions than listening. On the other hand, there was no difference on questions targeting literal comprehension. Further, reading was more beneficial than listening when there was no time limit imposed by the experimenter. Although Clinton found differences in reading and listening comprehension as a function of level of comprehension (inferential versus literal) and pace (self-paced or experimented-paced), the effect sizes were small. The meta-analytical review reflected the state of the research, which lacks a nuanced investigation of *for whom* and *under what conditions* comprehension with print and audio converge or diverge.

Motivated by the apparent lack of a deep analysis of the state of research, Singh and Alexander (2022) undertook a systematic review of the literature spanning 1970 to 2020. They posed the following critical question to guide their review:

What patterns can be uncovered about comprehension when the medium is an auditory versus written text and what characteristics of text, the learner, and the context, or the comprehension outcomes affect the patterns identified? (p. 3)

The research synthesis revealed that audio presentation supported comprehension as well as or better than its print counterpart when the text was narrative. However, with expository texts, print proved to be superior to audio. With mixed genre texts (i.e., containing elements of both narratives and exposition, Alexander & Jetton, 2000), audio

and print led to similar comprehension outcomes, but this was based on the results from only two studies. Audio co-presented with text (commonly called *reading-while-listening*) appeared to be better suited than reading by itself for EFLs and other English language learners. The authors noted that the state of the research did not afford further definitive analyses to derive conclusive claims. Therefore, they turned their attention to critically reviewing the studies and identifying gaps that can inform future research.

Gaps in the Research

Singh and Alexander (2022) highlighted several gaps in the literature that impede what we know about comprehension across print and audio mediums. These gaps were related to learner characteristics, text features, assessments of comprehension, the design, and the analysis conducted in the studies.

Learner Characteristics

The authors found that college students were underrepresented in the research even though college-educated people constitute the primary demographic listening to audiobooks (Pew Research Center, 2019). Of the 32 identified studies, only five included college students or adult participants, while elementary school-aged children were the primary population of interest. Further, most studies that included adults were conducted with English as Foreign Language (EFL) learners, and only two of the studies had native speakers. The emphasis on EFL learners is unsurprising given the importance of listening skills in courses teaching foreign languages (Dunkel, 1991; Goh, 2000; Zhang, 2013). However, it leaves a gaping hole in our understanding of the effectiveness of the audio medium when compared to print in supporting comprehension for college students who are reportedly a major user group of audiobooks. College students also process text of

varying complexity and genres and read for multiple purposes. Therefore, there are long-standing calls to increase text processing research with this subgroup (Alexander, 2005).

In addition to the limited coverage of the age groups studied, the systematic review also showed that affective learner characteristics had not been given their proper due in the research. Although nearly one-third of the studies included some measure targeting motivation, perception, or attitudes of learners, none of the investigations probed the role of these affective variables in influencing comprehension. For example, Nakashima and colleagues (2018) tested Japanese EFL students' comprehension in three conditions of text processing—reading, listening, and reading while listening. They also asked students to report their perceived difficulty and preference for the above three conditions through free-response questions. Students' responses indicated that listening was perceived to be the most difficult while preference for the three conditions was equally distributed. Although Nakashima et al. collected perceived difficulty and preference data, they did not relate it to comprehension outcomes. Given that a rich tradition of empirical research has established the intertwined nature of motivational constructs and comprehension (Guthrie & Wigfield, 2000), there is a need to uncover the influence of affective learner characteristics on comprehension across the audio and print mediums.

Text Features

Another significant finding reported by Singh and Alexander (2022) was related to text features. Most studies reviewed included the narrative genre (n = 20). Expository texts were included in ten studies, while only two had mixed text genre. The synthesis of the studies included in the review revealed that narratives appear to be better suited for

delivery via the audio medium than expositions when the goal is comprehension. The familiar story structure or arc that typifies a narrative could explain the suitability of the audio medium for this genre. On the other hand, expositions lack a predictable structure. Therefore, there is a higher reliance on textual elements such as headings and subheadings (Rubery, 2011; Surber & Schroeder, 2007). These textual elements are more salient in print than in audio. Due to a paucity of investigations with mixed genre texts, it is unclear if the narrative component of the text can compensate and lead to comparable comprehension as print.

Test Features

In terms of measurement of comprehension, the authors found that certain test formats were more favored than others and that most researchers tended to target comprehension at one specific grain size. In the 32 studies analyzed, comprehension was most commonly assessed with researcher-developed measures (n = 23), and most studies did not report the psychometric properties of the tests. Moreover, the researcher-developed comprehension measures tended to consist only one of the following formats—multiple-choice items (n = 8); gap-filling exercises including cloze and maze (n = 4); short answers or Wh-questions (n = 3); free and cued recall (n = 4); and matching (n = 3). Several studies included no information on test format but noted the level of comprehension assessed (e.g., Moyer, 2011: literal and inferential).

In nearly one-third of the studies (n = 9), vocabulary was the comprehension outcome of interest. Vocabulary gains were measured through various standardized measures, including the Peabody Picture Vocabulary Test (e.g., Boeglin-Quintana & Donovan, 2013) and the Listening Vocabulary Levels Test (Milliner, 2019). In addition,

for studies with EFL learners, researcher-developed vocabulary tests were administered that included multiple-choice and matching items presented in two languages (Brown et al., 2008; Webb & Chang, 2015). For example, Brown et al. (2008) embedded nonsense words in an English story that Japanese college students processed either by reading only or by listening and reading simultaneously (co-presentation). Students were then asked to select the correct meaning of the nonsense words from the four options that were presented in Japanese. Singh and Alexander (2022) found that *how* text comprehension was measured (test format) and *what* was measured (e.g., vocabulary, main idea) impacted the findings (Jenkins, 1979). In the studies they reviewed, the comprehension measures did not reflect the multidimensionality and complexity of the construct.

Research Design and Analysis

The majority of studies reviewed by Singh and Alexander (2022) employed a between-participants design. They all included variable-centered, quantitative analyses using statistical methods such as Analysis of Variance (ANOVA) tests. Through the synthesis of the research, the authors concluded that extending the research through person-centered analyses would allow researchers to pose more nuanced questions about the characteristics of learners who benefit more or less from listening to or reading texts. Moreover, to enhance our understanding of how processing behaviors and affect vary across groups of students for different mediums, it was beneficial to *complement* the quantitative findings with a rich qualitative analysis (Greene et al., 1989).

Therefore, I conducted a sequential mixed methods study using quantitative and qualitative methods to enrich our understanding of differences in comprehension across the print and audio mediums. The purpose of collecting and analyzing quantitative and

qualitative data was what has been labeled "*complementarity."* In such studies, quantitative and qualitative methods complement each other by measuring and explaining distinct aspects of the same phenomena (Caracelli & Greene, 1993). *Complementarity* is a key principle in mixed methods research wherein quantitative and qualitative data are inextricably linked together to provide a more complete understanding of a research topic. In this study, complementarity was achieved by using quantitative and qualitative data to form a more comprehensive understanding of the reader and listener profiles. The quantitative data were used to identify the profiles, while the qualitative data were used to further delineate and add depth to the nature of those populating the profiles.

Quantitative data were used to identify the reader and listener profiles through the use of finite mixture modeling. This approach allowed me to identify subgroups of readers and listeners based on patterns of behavior that were not apparent from the raw data. These quantitatively identified profiles were then used as a starting point for the qualitative data collection, which involved interviewing students from each profile group. The data from these interviews were analyzed using content analysis, which led to the creation of meaningful reader and listener profile groups.

The inclusion of qualitative data in this study helped to provide additional context and depth to the reader and listener profiles, resulting in a more complete understanding of these profiles. It was clear that delineating these profiles would not have been possible without considering the findings from both the quantitative and qualitative strands of the study together. The integration of both types of data allowed for a more nuanced and comprehensive characterization of the profiles and provided valuable insights into the behaviors and attitudes of readers and listeners.

The Current Study

Purpose

Given the documented rise in the use of audio as a delivery medium for text and the significant gaps unearthed by Singh and Alexander (2022) in the literature about learner characteristics, genre, measurement of comprehension, design and analysis, the purpose of this dissertation study was to expand the research on comprehension across audio and print mediums by employing a lens that focused on the learners. To achieve greater clarity around the critical question of mediums and student understanding, I conducted a *sequential explanatory mixed methods* study with a person-centered quantitative approach followed by cued retrospective interviews (Creswell et al. 2003; van Gog et al., 2005). In the first quantitative phase of the study, I collected data on learner characteristics, text and test processing behaviors, and comprehension indicators to form reader and listener profiles. For the second qualitative phase of the study, representative cases were selected for qualitative data collection using interviews from the profiles unearthed in the quantitative phase. The qualitative interview data were coded for analysis based on the same conceptual framework that drove the quantitative part of the study. Thus, the qualitative findings expanded and clarified the descriptions of the profiles derived from the quantitative analysis.

The population of interest for the dissertation study was undergraduate students. I used excerpts from a mixed genre text by Steven Pinker titled *Enlightenment Now*, which students processed visually and aurally. In this book, Steven Pinker, a cognitive scientist, uses data to argue that human wellbeing has improved over time. Each chapter of the book covers one indicator of human wellbeing. I chose excerpts from the chapters on

Inequality and *Happiness.* The complexity of the content in this text was suitable for a research study involving undergraduates since they are advanced readers required to read a variety of texts.

For the quantitative, person-centered part of the study, I conducted finite mixture modeling to identify clusters or profiles of readers and listeners (Vermunt & Magidson, 2002; Wolfe, 1970). The assumption underlying this statistical analysis is that the population distribution is a *mixture* of component distributions. Profiles are derived by identifying the unimodal component distributions that make up the complex multi-modal mixture distribution. In this study learner characteristics and their text and test processing behaviors were used as profile indicators. Learner characteristics included cognitive and affective variables. Specifically, I collected data on prior topic knowledge, self-efficacy in reading and listening, and reading and listening habits. In addition to these learner-related variables, text processing behavioral task-oriented data, such as scrolling, cursor movements, gazing patterns, re-listening, reading and listening time were also gathered.

Further, I also collected data on test related processing behaviors and variables, such as revising or re-framing an answer, and calibration of performance. The predictive validity of the identified profiles was determined using scores on the comprehension measures that targeted vocabulary gains, main idea, recall of explicit information, and inference. And prior topic knowledge was used in a covariate analysis to predict profile membership.

Through cued retrospective interviews, I gathered rich data on reader and listener profiles as they processed texts. This qualitative part of the investigation provided additional and clarifying insights into the nature of the reader and listener profiles, which

were inaccessible through a purely quantitative lens. These insights were integrated into the findings from the quantitative analysis to develop a nuanced picture of different reader and listener profile groups.

Research Questions

The guiding research aim of this study was to *identify and elaborate on the nature of different reader and learner profiles*. To fulfill this goal, I conducted a mixed methods investigation where the results from the quantitative analysis dictated data selection for the qualitative phase. The qualitative results were then combined with the quantitative person-centered analysis results to build an enriched understanding of reader and listener profiles underscoring the complementarity purpose of the mixed methods design (Clark, 2019). Before undertaking this mixed methods investigation, I conducted variable-centered analyses to determine the difference between print and audio mediums. This helped contextualize the current study vis-à-vis findings from previous research, which has been primarily variable-centered.

The following quantitative and qualitative research questions were posed. The findings from the qualitative phase further clarified and explained the different profiles unearthed in the quantitative phase, thereby underscoring the complementarity purpose of the mixed methods design (Caracelli & Greene, 1993). Due to the novelty of this research endeavor, it was not possible to forward hypotheses, and therefore, the questions were exploratory in nature.

Variable-Centered Descriptive Research Question

Medium Differences

1. *What are the differences in learner characteristics, text and test processing behaviors, calibration of performance, and comprehension performance between the print and audio mediums?*

Quantitative Research Question

Profile Identification

2a. What verifiable reader and listener profiles can be identified based on learner characteristics (e.g., self-efficacy, reading and listening habits), text and test processing behaviors?

Profile Validation

2b. How do students with different profiles vary on comprehension outcomes and calibration of performance between the print and audio mediums?

2c. To what extent does prior topic knowledge predict profile membership?

Mixed Methods Research Questions

Elaborating Medium Differences

3a. What are the differences between the print and audio mediums vis-à-vis learners, text, task, and test as reported in the interviews (qualitative results) and how do they relate to quantitative medium differences?

Elaborating Profiles

3b. What can the expressed ideas of the interviewees (qualitative results) elaborate and explain in the reader and listener profiles (quantitative results)?

Conceptual Model

The conceptual model for this dissertation research study was derived from Jenkins's (1977, 1979) tetrahedral model of memory experiments (see Figure 1). The

structure of the model offered the conceptual basis for understanding, designing, and analyzing the research problem at hand. It must be noted that this model served as the organizing heuristic for all aspects of this research study and was not a predictive model.

Recently, McMaster and Kendeou's (2023) articulated the heightened need for text comprehension researchers to refocus on theories that pay heed to the dynamic interrelations among leaner characteristics, texts, and their contexts. Likewise, Singh and Alexander (2022) also called for investigations comparing print and audio text comprehension that consider the characteristics of the learner, the text, the task, and the test, thereby accounting for the multidimensional nature of comprehension (Snow, 2002). Jenkins's model which he called the "problem pyramid" and "theorist's tetrahedron" provides the ideal framework for conceptualizing such an investigation of text comprehension. The assumption underlying the model is that comprehension is a complex and multidimensional construct. The three-dimensional structure reflects the complex interactions among different groups of variables known to influence comprehension (Cain et al., 2001; Kendeou et al., 2009; Kintsch, 1988). As Figure 1 depicts, each tetrahedron vertex represents one group of variables corresponding to learners, text, task, and test. Comprehension lies at the intersection of the planes.

The structure of the model suggests that comprehension results from the interaction of a *learner* who is doing the comprehending, the *text* that is being comprehended, the *task* where the comprehension is taking place, and the *test* where comprehension is externalized (Pearson & Cervetti, 2015; Snow, 2002). Each vertex influences the extent and nature of the meaning that a reader or listener extracts and constructs from the text. The specific variables considered in this dissertation study are

provided for each vertex, but it must be noted that many other variables that influence comprehension exist as well.

The model's structure also informs the organization and content of the subsequent chapters. Central to Chapter II (Review of Relevant Literature) as to this model is the nature of comprehension and its interrelations with all the variables related to the learner, text, task, and test. Chapter III offers the clearest picture of how the conceptual model has informed all aspects of the study, such as the operationalization of the constructs and resultant measures, the manner of data collection, the coding of the qualitative interviews, and the data analysis.

Key Terms

Text

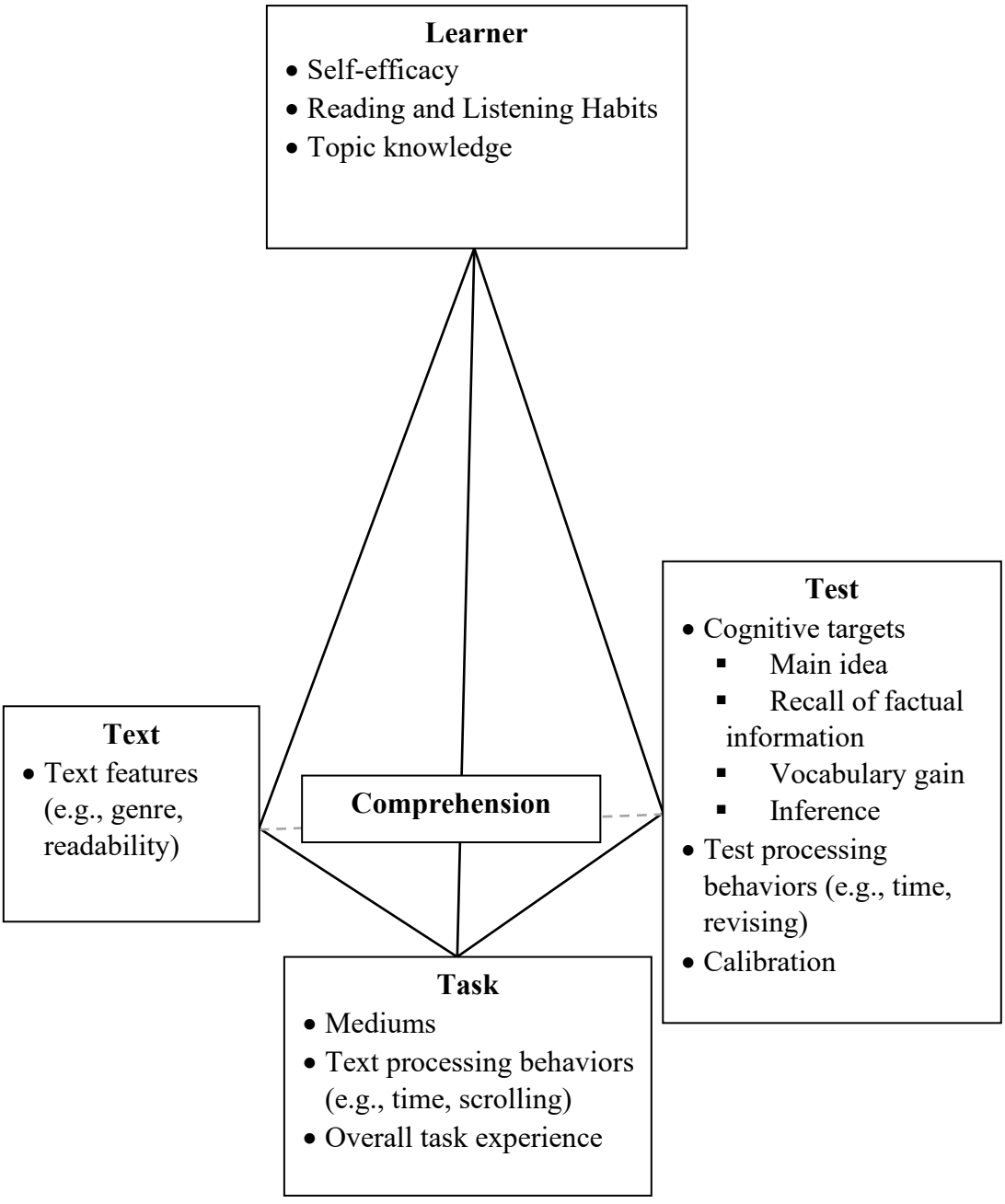
For the purpose of this study, text was defined as connected discourse containing more than a few words or phrases to allow for meaningful assessment of comprehension. This decision led to the exclusion of isolated words, brief signage, or slogans that have been used in some previous research studies comparing listening versus reading (e.g., Furnham et al., 1987).

Text Comprehension

Comprehension is a multidimensional process whereby readers or listeners construct meaning through their interactions with text (Fox & Alexander, 2009; Snow, 2002). This process relies on the interplay of cognitive, contextual, and affective processes (Alexander & The Disciplined Reading and Learning Research Laboratory [DRLRL], 2012). The meaning-making process may or may not result in an external product.

Figure 1

Conceptual Model for the Study



Medium

The definition of medium adopted in this proposed study is an instrumental one wherein medium is seen as a conduit for text delivery. The medium conditions under investigation will be text delivered through *audio* and *digital screen* (hereafter referred to as *print*).

Mixed Genre

A category of text that includes the characteristics of expositions and narrations. They contain elements of both expositions (e.g., facts) and narratives (e.g., dialogues). In effect, texts belonging to the mixed genre are dual purposed—they inform and entertain (Alexander & Jetton, 2000).

Strategies

Strategies refer to purposeful actions taken by a reader or listener to comprehend text (Alexander et al., 1998; Olshavsky, 1979). Within this dissertation, students' strategy use was investigated through cued retrospective interviews.

Self-Efficacy

For the purpose of this study, self-efficacy was limited to the domains of reading and listening. Therefore, referred to a student's specific beliefs in their ability to comprehend texts delivered through the print and audio mediums (Bandura, 1997).

Prior Topic Knowledge

A critical variable influencing comprehension, prior topic knowledge, indicates the individuals' pre-existing understanding of the text subject (Alexander et al., 1994). In this dissertation, prior topic knowledge was measured through specialized vocabulary drawn from the excerpts.

Calibration of Performance

A metacognitive construct, calibration of performance, is defined as the difference between actual performance and the student's own judgment of performance.

Text Processing Behaviors

Specific actions taken by individuals while interacting with text presented via print or audio usually with the purpose to derive meaning from the text (on-task behaviors). These actions are further divided into on-task and off-task behaviors. Example on-task behaviors coded were scrolling, forwarding audio. Sample off-task behaviors included eating or drinking, fidgeting.

Test Processing Behaviors

In the context of this dissertation, test processing behaviors refer to various ways in which students may approach or engage with the comprehension tests. Specifically, I coded for frequency of revisions, reframing constructed responses, non-sequential answering, that is returning to a prior question in the set, and scanning the test before beginning to answer.

CHAPTER II: REVIEW OF RELEVANT LITERATURE

To model "comprehension" in all its complexity is either impossible or very hard to do. But what one can do is decompose the problem into a set of reasonably independent components that can be studied in isolation, and that can then be combined to evaluate their interactions.

Walter Kintsch (1979, p. 4)

This chapter provides an overview of the relevant literature that informed this dissertation study. The chapter is purposefully laid out to parallel the conceptual model structure introduced in Chapter 1. Central to this chapter, as also to the conceptual model, is the multidimensional construct of comprehension. Therefore, starting at the intersection of the tetrahedron, I first cover the theory and empirical research on *text comprehension* highlighting how it is defined, particularly in the context of older and more experienced learners. Further, I discuss how comprehension has been measured, the different dimensions that constitute this construct. Second, I move to the *task* vertex of the tetrahedron and discuss the pertinent empirical work with mediums and their relation to comprehension.

Third, based on the dictates of the model, characteristics related to *learners* critical to comprehension are explicated. Specifically, I explore the literature on learner characteristics, such as prior knowledge, self-efficacy, habits and preferences, and what we know about learners' judgment of their comprehension performance (calibration of performance). The final section of the chapter covers behaviors exhibited during the text processing (e.g., re-reading or re-listening) and test processing (e.g., revising). The goal of this purposive survey of diverse research related to comprehension is to lay the

theoretical foundation for building the connections between variables that inform the proposed study and to articulate a case for undertaking this investigation.

Comprehension

Nature of Comprehension

Comprehension is the process of constructing meaning. It results in the formation of mental representations that may or may not be externalized (Kintsch, 1986). Comprehension is a multidimensional, multicomponent cognitive process whereby readers or listeners construct meaning through their interactions with text (Fox & Alexander, 2009; Snow, 2002). Those person-text interactions occur for diverse purposes (e.g., completing an assigned task, gathering information, gaining knowledge, or personal enjoyment) and in different mediums (e.g., paper, screens, or audio). Yet, for whatever purpose and in whichever medium, comprehension remains a complex activity that unfolds in a specific context and hinges on the abilities, skills, and behavioral processes of learners (Alexander & the Disciplined Reading and Learning Research Laboratory [DRLRL], 2012; Guthrie & Wigfield, 2000; Rumelhart, 1994).

With its focus on both the product and process of comprehension, the above definition is a result of what van den Broek and colleagues (1999) identify as the culmination of two generations of reading research. In the first generation, researchers focused on the mental representations or the *product* of interactions with text. This line of research aimed to understand the nature of the resultant mental representations. Some of the influential work that arose at the time looked at the influence of text structures on mental representations, such as story grammars for narrative texts and hierarchy of content organization for expository texts (Mandler & Johnson, 1977; Meyer, 1975). With

the development in experimental methodologies, such as eye-tracking and probing techniques, the emphasis shifted to the *process* of comprehension as researchers turned attention to online behaviors in the 1980s. This research aimed to examine what learners do as they proceed through the text (Fletcher & Bloom, 1988; van den Broek, 1990). Kintsch (1988) developed the influential construction-integration model of comprehension at this time. The construction-integration model highlighted the importance of prior knowledge and drew attention to the connections learners form between the ideas encountered in the text and their background knowledge.

Construction-Integration Model of Comprehension

According to the Construction-Integration model, comprehension processes are similar whether the textual material is listened to or read (Kintsch & van Dijk, 1978). Therefore, it has been used as the theoretical basis for studies in reading and listening comprehension (e.g., Clinton et al., 2020; Kendeou & O'Brien, 2018; Kim & Petscher, 2021). The model delineates three levels of representations required for constructing meaning from text—surface structure, textbase, and situation model (Kintsch, 2018; Kintsch & van Dijk, 1978).

The surface structure can be thought of as meaning elements that need to be extracted from the text. The literal words (semantic information) and phrases (syntactic structures) are examples of the meaning elements. In print, the receptacle for this information is written words, whereas in audio the meaning elements are carried in tone and prosody. In the context of reading, decoding is a necessary skill for this level, and for listening, it is the ability to recognize the sounds of words. The emphasis on decoding coincides with the Simple View of Reading (SVR) where comprehension is viewed as

product of two components—decoding and listening comprehension (Hoover & Gough, 1990). Thus, the SVR establishes the link between reading and listening comprehension. However, for the purposes of this study with advanced readers (college students), it is imperative to consider other variables and processes.

The second level, textbase, uses the surface structure information to develop sentence and discourse-level structure (*macrostructure*). The discourse-level structure is predicated on the global organization of the text. For instance, narratives usually follow the setting-complication-resolution structure. In contrast, expositions are organized according to various rhetorical formats, such as compare-and-contrast format or giving a general statement followed by examples. This aspect of the Construction-Integration model has implications for genre's influence on comprehension and if certain mediums may be better suited for certain genre types.

The third and highest level of representation is the *situation model*. The situation model results from integrating the information contained in the text with the learner's prior knowledge. Thus, prior knowledge is seen to play an integral role in the process of comprehension. This is important for the proposed investigation where the interrelations among text and learner characteristics (e.g., prior knowledge) will be studied.

Levels of Comprehension

Foundational to the assessment of comprehension is the theoretical definition of this construct. In the definition presented above, comprehension is viewed as a complex process comprising multiple components. In response to the complexity of the activity and the diversity of elements encompassed in text comprehension, the approaches to its assessment are similarly complex and varied. The assessments target different levels of

comprehension arising from varying levels of cognitive processing through different test and items formats.

A brief review of the literature reveals that although the phrase “levels of comprehension” is widely used, it is rarely defined and varies on what the levels encompass. Studies in the field of psycholinguistics begin the study of different levels from word recognition graduating to meaning retrieval, and then moving on to syntax. Other scholars frequently refer to only two levels: literal and inferential (e.g., Bishop & Adams, 1992; Diakidoy at al., 2005). And yet another group of researchers slices up the levels further (Bernstein, 1955; Schiefele, 1992). For example, Schiefele (1992) does not define the levels or the skills underlying them but lists the following skills in ascending order of complexity: retention of facts, recognition of relationships, coherence of text content, integration with prior knowledge, and recognition of main ideas.

The commonality in the varied ways that levels of comprehension are operationalized is the idea that comprehension requires the enactment of increasingly complex cognitive processes (Gernbacher et al., 1990; McNamara & Magliano, 2009). van Dijk and Kintsch (1982) provide one of the few theoretical descriptions of the levels of text comprehension, which serves as a valuable guide for constructing assessments. They delineate three levels: (a) the surface form; (b) textbase; and (c) situation model. Surface form and textbase comprise lower-level comprehension processes, such as accessing the meaning of words, sentences, or phrases.

The situation model or building mental representations of the text relies on higher-level comprehension processes that integrate text-level understanding with prior knowledge. These levels have been empirically tested. Past research has used recall of

concrete facts explicitly stated in the text to measure lower levels of comprehension and inference and gist understanding questions to target higher levels of comprehension (Dunlosky & Rawson, 2005; Haenggi & Perfetti, 1992). The levels of comprehension are also reflected in NAEP's notion of *cognitive targets,* which are used to develop different types of questions for the reading assessments (National Assessment Governing Board, 2015). The NAEP framework has identified three cognitive targets operationalized in the following kinds of questions—locate and recall, integrate and interpret, and critique and evaluate.

Summary

The definition of comprehension, the brief description of the foundational theory, and the levels of comprehension had implications for this study's design. First, the process, the product, and their complex interrelations need to be studied to better understand and describe comprehension across print and audio mediums. Therefore, I employed a mixed methods approach where the person-centered quantitative and qualitative analysis worked together to present a fine-grained picture of the complex nature of comprehension across mediums. Second, comprehension hinges on a slew of variables related to the learner (e.g., prior knowledge, self-efficacy), text (e.g., genre), processing behaviors that might be medium-dependent, and the nature of the measurement. With this in mind, in this study, I collect data related to: (a) learner characteristics; (b) processing behaviors while learners engage with the text and afterward when they take the comprehension tests; and (c) measured comprehension at different levels through selected and constructed response items.

Medium of Text Delivery

Theoretical examinations of the notion of medium are abounding because of the importance of medium for disciplines such as media studies, literature, human-computer interaction, and educational psychology (Freeman, 2016; Haas, 1989; Mangen & van der Weel, 2016; McLuhan, 1964;). For this study, the definition of medium is centered on two aspects—function and features. I regard medium first and foremost as a delivery channel for text. Second, given that this investigation focuses on comparing the "new" audio and screen mediums, it is necessary to attend to the features and affordances of mediums that determine how we interact with them differently (Have & Pedersen, 2016; Mangen & van der Weel, 2016). For example, readers can scroll through text while listeners can re-listen to text by rewinding. It is possible that scrolling on the screen might be more frequent than rewinding (or forwarding) audio, and this could influence comprehension levels. It is questions such as this that I explored in this dissertation. Previous research comparing print and audio, and paper versus screen provided some fruitful lines of inquiry and showed the gaps in the existing research informed this study.

Print versus Audio Mediums

Reviews of the Literature

Given the popularity of the audio medium in the 1960s and 1970s, it is understandable that there have been reviews investigating comprehension differences when reading or listening from that era (Duker, 1965; Sticht et al., 1974). Duker (1965) synthesized findings from 34 studies and concluded that print and audio mediums supported comprehension equally. However, audio appeared to be better suited for younger students than older ones who performed better with reading. Further, Duker

noted that reading appeared to support learning from text, whereas listening was better for retaining details.

Sticht and colleagues (1974) investigated the viability of "auding" as a pedagogical tool for supporting literacy in young military recruits by conducting a comprehensive review of research on listening and reading. They determined that younger students (until grades 6-7) are better at listening than reading. However, the reverse is true when students get older, with reading skills having an advantage over listening skills. The researchers used the findings from the review to conclude that training recruits in listening comprehension would transfer to reading comprehension.

Much has changed in terms of the mediums, students, and pedagogy since the 1970s. Advancements in mobile, file storage, and recording technology have revolutionized access to the variety of and quality of audio textual content (e.g., audiobooks, podcasts). Students are exposed to diverse mediums. The Common Core Standards state that students should be able to evaluate and integrate information presented orally and visually (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010).

Given these changes, two reviews were carried out recently. The first, a meta-analysis by Clinton-Lisell (2021), investigated reading and listening comprehension comparison studies across different age groups. The review included 36 reports with 46 independent samples from 1966 to 2020. The participants ranged from elementary-school-aged children to college-level adults. Clinton-Lisell provided information on the genre of the texts used (narrative, expository, mixed) and also noted the texts' languages (e.g., English, German, Greek). Despite the more naturalistic nature of the studies

compared with the neuroscience research, the meta-analysis also found no statistically significant difference in reading and listening comprehension similar to some of the findings from neuroscience studies (e.g., Deniz et al., 2019). The next step involved moderator analyses to determine if any variables may result in differences between reading and listening comprehension. Surprisingly, there were no reliable differences in reading and listening comprehension as a function of genre, age group, orthography, and study design. The only significant moderators were assessment type and pacing. Reading was more beneficial than listening for answering inferential questions and when the reading condition was self-paced.

The second, a systematic review by Singh and Alexander (2022), purposefully analyzed the literature to determine for whom and under what conditions comprehension differs across print and audio mediums. The authors reviewed 32 empirical investigations and found that co-presenting print and audio is more beneficial than reading by itself for EFLs and other ELLs. They also found that audio is as effective or better than print only for narrative or mixed genre texts. There was some indication that listening was more advantageous for younger children. In addition to the findings, the strength of this review was in the gaps it uncovered in the literature. The following gaps are highlighted because they had a direct bearing on the design of the study: (a) there is limited research comparing print reading by itself to audio listening; (b) most studies target younger populations; and (c) fewer studies focus on expository or mixed texts. The authors' primary goal in the review was to determine what characteristics of the learner, the text, the context, and assessments affect comprehension across the print and audio mediums. However, they noted that the state of the research is not sufficiently developed to provide

such a nuanced understanding. Therefore, the proposed study investigates the interrelations between learner, text, and test dimensions and their influence on comprehension across mediums.

Neuroscience Research

Brain imaging research comparing reading and listening comprehension has mixed findings. Several studies indicate that reading or listening to texts activates similar brain regions (e.g., Braze et al., 2011; Deniz et al., 2019). In the study conducted by Deniz and colleagues (2018), adult participants listened to and read the same stories while their brains were scanned using functional Magnetic Resonance Imaging (fMRI). They found that the same cortical regions associated with processing semantic information were activated in the reading and listening conditions. Although the study may have found that comprehension processes may be medium-invariant, it is essential to note that in the reading condition, each word was presented on the screen one at a time (rapid serial visual presentation). This was done so that the researchers could determine precisely what word a reader was reading. The reading context was not naturalistic or ecologically valid. However, there is no denying that shared linguistic subskills, such as vocabulary knowledge and inference making, are equally crucial for reading and listening to verbal texts, which might explain the finding from brain imaging studies (Kendeou et al., 2008; Wolf et al., 2019).

Other researchers have found medium-specific brain activation associated with reading and listening (e.g., Buchweitz et al., 2009; Constable et al., 2004). Buchweitz and colleagues directed adult participants (enrolled or graduated from college) to read and listen to 12 sentences in each condition in an fMRI machine. The visual stimuli were

presented one word at a time, similar to the protocol followed by Deniz and colleagues. The words were spoken with little to no prosody, and the sentences did not contain any punctuation marks other than a period. Both visual and auditory stimuli had to be sequentially processed so that participants could not backtrack. The researchers found common and medium-specific brain activation regions. They also reported observing individual-level differences in this imaging study.

Currently, the state of technology is constraining the type of questions that can be posed and, consequently, the conclusions drawn from neuroscience research. The constraints are reflected in the simplistic nature of stimuli, the delivery mediums, and the experimental conditions that are too far removed from the usage context. As a result, the studies are not ecologically valid, making the call for behavioral investigations stronger where the interactions between variables known to influence comprehension can be studied under different medium conditions.

Paper versus Screen Mediums

While the state of the research comparing print and audio mediums is wanting in terms of the complexity of the investigations, there is now well-developed literature comparing comprehension across the print and digital that offers insights into the type of questions that medium-research can address. Similar to the previous section, I present the reviews of research that have been undertaken so far and then discuss purposefully selected empirical studies that can guide the proposed dissertation study.

Reviews of the Literature

Given the continual advancements in medium-related research with paper versus screen, successive meta-analytic and systematic reviews have posed different questions to

the literature. In this section, I will cover three reviews that have each made distinct contributions to understanding how medium interacts with different variables to influence comprehension. All the research syntheses provide converging evidence that medium indeed plays a critical role (Clinton, 2019; Delgado et al., 2018; Singer & Alexander, 2017a).

The earliest of the three research syntheses was undertaken by Singer and Alexander (2017a). They assembled a pool of 36 empirical studies conducted between 2001 and 2017. They reported research trends related to age group of the learners, the genre of the texts, the features of the medium, and the assessment of comprehension. They found that research has been focusing on a wide range of school-aged children from elementary to college-level. The text genre used was tied to the age-group. Most studies with elementary school-aged children employed narratives, while college readers were tested with expositions. Importantly, the authors found that certain text and medium features were being manipulated. For example, several studies investigated the influence of scrolling on comprehension, and some studies manipulated the font size and sentence spacing in the digital conditions. It appears that in the paper versus screen reading research there is a greater acknowledgment of the affordances of medium than in the print versus audio research. In relation to assessment, Singer and Alexander found that more than half the reviewed studies contained multiple indicators of reading comprehension.

Delgado and colleagues (2018) conducted a meta-analysis of 54 studies with the goal to compare reading texts on paper and digital device. The reports were drawn from the same time period covered by Singer and Alexander (i.e., 2000 to 2017). Overall, the authors reported a comprehension advantage for reading on paper than reading digitally.

Further, their extensive moderator analyses yielded several crucial significant findings. First, reading on paper was more beneficial than digital reading irrespective of age group. Second, this paper advantage was found for short and long texts. Third, reading on paper yielded better comprehension outcomes than digital reading when the text was expository or mixed.

Finally, texts presented on paper were better comprehended than those presented digitally as evidenced by higher scores on textual, inferential, or mixed comprehension questions. Processing digital texts appeared to be as effective as paper reading on handheld devices, contexts where learners did not need to scroll, and where they could control the amount of time spent (self-paced condition). This meta-analysis underlined the influences that variables related to genre, affordances of the medium that lead to different types of text processing behaviors (e.g., scrolling, tactility), and the context of reading (e.g., pacing). These findings from Delgado et al. highlight that variables related to genre, text processing, and levels of comprehension need to be considered in medium-related research.

Clinton (2019) systematically reviewed and meta-analyzed 29 empirical studies to synthesize the findings on reading comprehension, reading times, and calibration of performance across paper and digital screen mediums. Similar to Delgado et al. (2018), Clinton found an overall advantage for reading on paper than reading on screen for comprehension. Further, comprehension findings did not vary by learners' age-group. Genre also emerged as a significant moderator with paper being more advantageous in the case of expository texts only. Corroborating the findings from Delgado et al.'s (2018) meta-analysis, Clinton showed that the benefit of paper versus screen was maintained for

literal and inferential comprehension questions. The unique contribution of Clinton's meta-analysis to the medium-research relates to reading times and calibration of performance. She reported that there was no consistency in reading times across mediums and called for further research in order to learn how medium influences reading time. There were several studies that had investigated calibration of performance across mediums. Students were more overconfident in their performance and poorly calibrated when text was read digitally than on paper. The findings from all three reviews highlight the variables that seem to interact with medium to influence comprehension. It is important to look at the studies themselves to understand specifically how these variables were manipulated and investigated.

Experimental Studies

There is a growing corpus of research on reading comprehension across different mediums (Daniel & Woody, 2010; Kendeou et al., 2005; Støle et al., 2020). These "paper versus digital" researchers have investigated differences in reading speed, accuracy, recall, and comprehension as a function of the medium (Ackerman & Goldsmith, 2011; Singer Trakhman et al., 2018). For instance, Ackerman and Goldsmith (2011) found that undergraduate students were much quicker in reading digitally than in paper, but worse at comprehension as measured by multiple-choice questions. More importantly, when those students were made to read on different mediums for the same amount of time, the difference in comprehension performance disappeared. These researchers attributed this difference in reading speed to variable levels of self-regulation afforded by the mediums.

Singer and Alexander (2017b) also found reading comprehension differences across the paper and digital mediums in their studies with college students. They asked

college students to rate their preferred medium of reading. They assessed comprehension through constructed-response questions that tapped into the main points, key details, and an open-ended question about any other details that students recalled. Despite a preference for reading digitally and self-assessments that rated comprehension performance higher for digital texts, students comprehended better in paper. In a follow-up study, Singer Trakhman et al. (2018) investigated the nature of comprehension differences across the two mediums. They found that students could recall the main idea of texts equally well across mediums, but details were recalled better when the text was accessed on paper. The researchers explored processing time as a potential mediator between medium and comprehension. They found that processing time could be a potential explanatory factor for comprehension differences across mediums. They also noted that the extra navigational effort required for digital reading (e.g., scrolling) and the absence of spatial anchoring could be contributing to these differences.

Summary

There is an extensive body of research comparing listening and reading comprehension spanning over 100 years (Sticht et al., 1974), which arguably raises the question of why another study comparing print and audio comprehension should be undertaken at this time. As established in this section, there are three key reasons for conducting the current study. The first relates to the type of participants. Previously, reading and listening comprehension has been studied with younger children learning to read and second and foreign language learners (Singh & Alexander, 2022). In contrast, in this study, undergraduate students who do not need any specialized language instruction were the population of interest.

The second reason relates to the research questions pursued in research conducted in the past. Much of the earlier work focused on uncovering the correlation between reading and listening comprehension in an attempt to understand the relation between these two competencies especially as a function of age (Clinton-Lisell, 2021; Verlaan et al., 2017). The empirical work emerging from this era established that listening comprehension is a strong predictor of reading comprehension development (Anderson et al., 1985; Sticht & James, 1984).

In the 1970s, other researchers focused on the cognitive processes presumed to underlie reading and listening comprehension and concluded that these two mediums relied on the same cognitive processes (Berger & Perfetti, 1977; Kintsch & Kozminsky, 1977). Indeed, the simple view of reading purports that reading comprehension can be explained simply as the ability to decode combined with listening comprehension ability (Gough & Tunmer, 1986; Hoover & Gough, 1990). In contrast to this unidimensional view of comprehension in this line of research, in the investigation, I provided a nuanced picture of the nature of differences in learners as they read and listen to texts to construct meaning. In other words, instead of a simple correlation between variables, I looked at the complex interrelations between learner characteristics, task, and test processing in students reading and listening to texts.

Third, the previous research studies did not consider the context of print and audio delivery through different mediums. The importance of the question of medium is pertinent in the present era of screen reading and audio listening. The past literature has not generally studied medium-related differences that might directly influence students' comprehension, as has occurred in the studies on paper versus digital reading (reviewed

in the subsequent section). For example, in those studies, medium features, such as demands of scrolling or treatment of multimedia have not been systematically analyzed (Mayer, 2005; Sanchez & Wiley, 2009). They did not consider the interrelation between the text, listener/reader, and the medium. For instance, speed of listening versus reading, the likelihood of re-listening and re-reading, quality of the recording, or genre of text were not addressed.

All the research syntheses provide converging evidence that medium indeed plays a critical role (Clinton, 2019; Delgado et al., 2018; Singer & Alexander, 2017a). This is in contrast to the print versus audio meta-analysis and neuroscience literature findings, wherein no reliable differences were found in comprehension across mediums. I argue that the similarity in comprehension reported in the print and audio literature is a consequence of the stunted nature of the investigations. The primary goal of this deep dive in paper versus screen literature was to draw out what we can learn and transfer from this research arena to its neighbor—the research world of comprehension across print and audio.

Learner Characteristics, Medium, and Comprehension

As established, an investigation related to medium and comprehension would be incomplete without considering the learners who are engaged in the acts of listening or reading. Learner characteristics such as age group, self-efficacy as readers and listeners, prior knowledge, reading and listening habits are all factors that influence comprehension and the nature of engagement with the text through different mediums.

Age-Group of the Learners

The participant's age is an essential factor to consider when studying text processing in any medium (Chall, 1983). This section compares reading and listening development and discusses studies investigating the relation between reading and listening comprehension. Further, I focus on the skills that underlie reading and listening and what it means to acquire competence in comprehending text, whether written or oral.

Reading and Listening Development

Reading is an acquired skill, and the route to comprehension starts with a word (Ehri, 2005). Beginner readers have to learn to recognize words and activate the meanings they represent. Successful word reading (in alphabetic languages) hinges on recognizing letters and accessing the corresponding sounds (phonemic awareness). Early readers, therefore, have to learn to sound out words, identify new words by noticing how they are analogous to already known words, and concurrently develop sight word reading which relies on recognizing words from memory. With sight word reading, word pronunciation and meaning is accessed automatically (Ehri, 1992; 2005; LaBerge & Samuels, 1974). These lower-level reading processes constitute the critical skill of decoding.

In contrast, most children acquire basic listening ability naturally. Children without hearing deficits learn to separate sounds of and identify words in all languages exposed to at an early age without specialized instruction. The difference between the nature of reading and listening is evident in the brain itself. Reading does not have a specific brain area. According to the neuronal recycling hypothesis, we repurposed a subset of the visual cortex, initially reserved for facial recognition, to recognize letters

(Dehaene, 2005; 2014). Once the written words are recognized, they are processed the same way as spoken words (Braze et al., 2011; Dehaene, 2020; Deniz et al., 2019).

Relation between Reading and Listening Comprehension with Age. Several studies have investigated the relation between reading and listening comprehension with mixed findings. Sticht and colleagues (1974) argued that listening comprehension skills are stronger than reading comprehension skills till grades sixth or seventh, after which reading comprehension takes the lead. In a more recent study comparing reading and listening across grade levels, Verlaan and colleagues (2017) found that reading comprehension was consistently better than listening comprehension at grades four, six, and eight. It was the size of the difference between the two types of comprehension that differed. There was a small, medium, and large difference between reading and listening comprehension at grades four, six, and eight, respectively.

Diakidoy and colleagues' (2005) study provided additional data about the interactions among grade level and comprehension by including the variable of genre. The researchers recruited students from grades two, four, six, and eight. Diakidoy et al. (2005) had students listen to and read both narrative and expository texts at each grade level. Their comprehension assessment was a sentence verification task that included 7 literal and 7 inferential statements that the students were directed to judge as true or not based on the text they just heard or read. The outcomes from this study offered an interesting developmental perspective on the interaction between genre and reading and listening comprehension.

Specifically, Diakidoy et al. (2005) found that listening comprehension was better than print at grade two but only for narrative texts. There was no difference in

comprehension for narrative texts at grades four or six, and by grade eight, reading narratives benefited comprehension significantly more than listening. For expository texts, listening and reading resulted in comparable performance in comprehension for students in grades two, four, and six. However, in grade eight, reading the printed text was significantly better for comprehension than listening. This study indicated that the properties of the text play a key role in influencing comprehension.

Singh and Alexander's (2022) systematic review of the literature found further evidence that audio was better suited for comprehension of narratives than expositions. They posited that the predictable structure of narratives made them more accessible through audio. They also noted that the reviewed studies did not provide information about the texts' linguistic complexity or readability levels. They forwarded the possibility that narrative texts used in the studies could have been at a lower readability level than the expository texts on average.

Proficient Reading and Listening. As students gain proficiency in emergent literacy skills, they transition from *learning to read* to *reading to learn.* Then the goal of instruction shifts to developing mature readers who can "read a variety of materials with ease and interest, can read for varying purposes, and can read with comprehension even when the material is neither easy to understand nor intrinsically interesting" (Snow, 2002, p.xiii). Alexander (2005) provides a framework for understanding the multiple factors that play a role in developing a competent and mature reader. Drawing on the Model of Domain Learning (Alexander, 1997, 2004), Alexander highlights the importance of cognitive and motivational factors in reading development beyond emergent literacy. Prior knowledge, interest in reading, and strategic processing are deemed critical for this

reading stage. These processes are deemed necessary for comprehending both oral and written texts (Alexander & Jetton, 2000).

For the purposes of the current study, where the focus was on processing a complex text across different mediums, mature and competent readers and listeners were the population of interest. College students represented this population, given that their academic path has made them adept at engaging with complex, information-heavy texts that may not be of interest. Such texts also rely significantly on the activation of prior knowledge, which presumably a college student possesses on a multitude of topics.

Prior Topic Knowledge

Prior topic knowledge refers to an individual's pre-existing knowledge of the text content. Prior knowledge plays a critical role in text processing and comprehension (Bartlett, 1932; Bransford & Johnson, 1972; Kintsch, 1988, 1998; McNamara et al., 1996). In Kintsch's Construction-Integration Model of comprehension, described in detail in the previous section, prior knowledge is theorized to be necessary because texts do not contain sufficient information for building a mental representation of the situation of the text. In response, individuals must supply the missing information for generating a coherent mental representation (i.e., situation model) from their personal store of knowledge (Kintsch & van Dijk, 1978). Empirical studies have corroborated the importance of prior knowledge to text comprehension, particularly in the case of expository texts (Afflerbach, 1986, Chi et al., 1981, Dochy et al., 1999; Haenggi & Perfetti, 1994).

In studying informationally-rich texts, researchers have focused on the interaction among prior knowledge and text features, such as coherence and structure (Kendeou &

van Den Broek, 2007; McNamara & Kintsch, 1996; Ozuru et al., 2009). Text cohesion is a function of the text's linguistic, conceptual, and organizational features. For example, a highly cohesive text contains more connective words, greater overlap in arguments between adjacent sentences, and the presence of headers and topic sentences (McNamara et al., 1996). A series of studies conducted by McNamara and colleagues demonstrated that readers with higher prior knowledge benefit from low-cohesion texts. In contrast, those with lower prior knowledge better comprehend highly cohesive texts. The observed differences in comprehension were also related to the level (e.g., literal or conceptual) and type of comprehension assessment (Dochy et al., 1999). In general, high coherence texts were related to better performance on recall and multiple-choice questions. However, those with higher prior knowledge performed better on sorting tasks that needed conceptual understanding after reading the low coherence texts (McNamara, 2001; McNamara & Kintsch, 1996). All these investigations focused on the product of reading by assessing differences in comprehension measures obtained after reading.

Kendeou and van Den Broek (2007) extended this line of inquiry on the interaction between prior knowledge and text features by examining text processing during reading. They used think-alouds and reading time to capture online cognitive processes as undergraduates read two types of science texts, refutational and non-refutational. The researchers assessed prior knowledge of Newtonian mechanics through the Force Concept Inventory (Hestenes et al., 1992). In addition to finding that the lower prior knowledge group produced lesser correct inferences during reading than the high prior knowledge group, they found that the type of text interacted with prior knowledge.

Specifically, those with lower prior knowledge generated more conceptual change strategies while reading refutational texts than non-refutational texts.

Although the research on the interrelations among learner characteristics, text features, and test characteristics is well developed, there is limited understanding of how different mediums interact with learners, texts, and tests. The research discussed above has established that text cohesion and structure is one crucial factor, and we know that different mediums afford access to text differently. For instance, texts with several elements (e.g., diagrams, figures, headings) are more suited for consumption through print than audio for the apparent reason that audio cannot encode images (Rubery, 2016; Surber & Schroeder, 2007). Further, texts enshrined in print have a quality of permanence that is lacking in audiobooks (Garner, 1987). Print affords readers the option to pause and reflect on the words or go back to text previously read with far greater ease than listening (Garner et al., 1984; Wittkower, 2011). Listening is more transient, increasing the challenge on listeners to re-play a previous segment of the text (Rubery, 2011; Leahy & Sweller, 2011). Indeed, learners who are in the acclimation stage, that is those who have limited prior knowledge and experience with expository texts, rely more on re-reading to construct meaning (Dinsmore et al., 2015). For such learners, the print medium that easily allows going back to the text would be more suited than audiobooks which do not afford easy re-access to text.

However, Singh and Alexander's (2022) systematic review of the audio and print literature did not find any studies that considered prior knowledge and its interactions with text and comprehension outcomes. And although the research comparing text comprehension on paper and screen (static text and videos) has included prior knowledge

as a variable based on the theoretical recommendations of the Construction-Integration Model (e.g., List & Ballenger, 2019; Singer & Alexander, 2016) it has not accounted for the possible variability in the influence of prior knowledge based on the medium of the text. For the purposes of the current study, the central role played by prior knowledge in comprehension across different types of texts and the evident absence of a nuanced consideration of prior knowledge in the medium-related comprehension research suggests that the investigation of the influence of prior knowledge must be taken into account.

Self-Efficacy

Bandura (1997) conceptualized the construct of self-efficacy as people's beliefs in their capabilities to organize and execute courses of action to attain a goal. This self-belief situated in social cognitive theory, Bandura explained, is domain-specific and not a global trait. In other words, individuals have varying levels of self-efficacy in different realms of functioning. For instance, a student may have a high level of self-efficacy in the domain of reading but low efficacy in the mathematics domain. Therefore, Bandura (2006) strongly cautioned against "all-purpose" measures of self-efficacy and emphasized the necessity to capture self-efficacy using measures specifically tailored to the domain of interest. Therefore, in this proposed study, I plan to adapt a reading self-efficacy measure also to capture listening self-efficacy beliefs. It is possible that reading and listening beliefs may co-vary. However, given the novelty of this area of research, I am unable to hypothesize the nature and degree of their relation.

Self-efficacy has been linked to achievement outcomes in different academic domains through multiple pathways. Self-efficacy influences choice, effort, goals, and persistence (Schunk & Dibenedetto, 2020). In the domain of reading, Shell and

colleagues (1989) investigated the relation between self-efficacy beliefs and reading achievement. Undergraduate students self-reported their perceived ability to execute tasks such as reading a 400-page novel or a scholarly article. Reading achievement was assessed using the Degrees of Reading Power test (DRP, Touchstone Applied Science Associates, 1983). DRP measures comprehension by giving participants passages with keywords missing. Participants are directed to select the most appropriate word consistent with the passage's meaning from multiple-choice options. Multiple regression analysis revealed that self-efficacy accounted for significant variability in reading achievement.

In addition to the relation unearthed between self-efficacy and reading achievement, another aspect of the study merits highlighting as it has implications for this dissertation research—specifically, the deliberate choice of undergraduates as the population of interest for this research question. Shell et al. (1989) chose undergraduates because they represent mature, skilled readers. The researchers discussed previous studies that showed the increasing influence of self-efficacy on reading achievement with increasing grade levels (Nicholls, 1979; Paris & Oka, 1986). They offered a theoretical explanation for this effect by grounding the findings in Bandura's assertion that the influence of self-efficacy increases with mastery of component skills (Bandura, 1986). For Shell and colleagues, this meant that the relative importance of self-efficacy on reading achievement would be greater for more skilled readers who have developed the cognitive and behavioral skills necessary for proficient reading. Indeed, the results from their study attested that self-efficacy explains a significant portion of the variability in reading achievement in undergraduates. This empirical finding and its theoretical backing

justify choosing self-efficacy as a key learner variable in this study with undergraduate students.

Reading and Listening Habits

Popular beliefs and evidence from research concur that exposure to reading activities is linked with better reading comprehension (Anderson et al., 1988; Cunningham & Stanovich, 1991; Guthrie et al., 1999, Mol & Bus, 2011). Frequent text exposure improves comprehension by building vocabulary (Krashen, 1989; Sénéchal et al., 2008), expanding content knowledge (Hirsh, 2003; Stanovich et al., 1995), and making reading more pleasurable, thereby leading to further motivation to engage with texts and increased achievement (termed "Matthew effect", Stanovich, 1986). The benefits associated with text exposure are not restricted to the act of reading. Indeed, text exposure via listening has also been related to improvement in vocabulary and other forms of comprehension (Penno et al., 2002; Swanson et al., 2011). While the research on reading and listening habits and their relation to comprehension is well-developed with younger populations, there is limited understanding of these relations with college students (Mol & Bus, 2011).

Mol and Bus (2011) addressed this gap by investigating how the correlation between text exposure and comprehension develops with age and emerges with older students. They meta-analyzed 99 studies, out of which 30 included undergraduate or graduate populations. They found a consistent association between text exposure and comprehension across grade levels. However, this finding was inconsistent with their hypothesis that predicted a stronger correlation between exposure and comprehension with increasing grade levels based on the model of reciprocal causation (Stanovich,

1986). The authors noted that the finding could have been a product of the variation in response formats in comprehension measures across grade levels. Multiple choice and open-ended questions, which may require higher-order skills such as integration, were primarily used in studies with college students. In contrast, relatively easier formats, such as cloze tasks, which rely on word reading abilities and sentence comprehension, were employed with the lower grade levels. In effect, the measures across grade levels were not targeting the same latent skills. Nevertheless, Mol and Bus's meta-analysis established that text exposure is as critical, if not more, in the case of college students for comprehension as it is for younger populations.

Although there is comprehensive coverage of age groups in this area, the research establishing the importance of text exposure to comprehension has been primarily conducted with traditional mediums of text delivery (McGeown et al., 2015). Given that students increasingly spend more time engaging in digital text activities, there is a need to broaden this field of inquiry to include other mediums like audio and reading on the internet. Indeed, in a survey of 894 college students' reading and listening habits, I found that 64% read digitally compared with 29% reading on paper. Further, 47% reported listening to audiobooks and podcasts for pleasure and 26% for school (Singh & Alexander, 2021). This highlights a shift in students' literacy habits. Although the reduction in paper-based book reading is often framed as a cause for concern (Huang et al., 2014; Mokhtari et al., 2011), research has found that the skills needed to read in digital contexts like the internet are relatively complex, comprising of inferential reasoning, prior knowledge, self-regulated processes (Coiro & Dobbler, 2007; List &

Alexander, 2017). Therefore, it was crucial to account for different literacy experiences in this study.

Calibration of Performance

The construct of calibration is derived from the study of metacognitive monitoring (Nietfeld et al., 2006; Schraw, 2009; Zimmerman, 2000). At its simplest, it reflects the relation between an individual's perceived performance and actual performance (Glenberg & Epstein, 1985; Lichtenstein et al., 1982). If a student feels that they performed poorly and correspondingly demonstrate low scores on the measure, they would be well-calibrated. Within the context of text comprehension, calibration is frequently captured by asking students to rate their performance after taking a comprehension measure and comparing that rating to the actual performance score (Dinsmore et al., 2015; Schraw et al., 1993).

Despite being extensively studied, calibration of performance remains one of those constructs whose theoretical underpinnings are not explicated in studies (Parkinson & Dinsmore, 2010). Dinsmore and Parkinson (2013) provided a helpful framework for understanding calibration. Recognizing that a host of factors influences calibration, they used Bandura's (1986) model of reciprocal determinism as a means to unpack those factors. Calibration within this model is influenced by personal, environmental, and behavioral factors. Personal factors, such as prior knowledge, have been theoretically and empirically shown to be related to calibration (Alexander, 1997; Glenberg et al., 1987). Environmental factors, such as item difficulty and type of feedback, also influence calibration accuracy (Ackerman & Goldsmith, 2011; Gutierrez & Schraw, 2014). Behavior, in Bandura's model of reciprocal determinism, is divided into three types,

namely, self-observation, self-judgment, and self-reaction. Calibration can be viewed as self-observation or self-judgment depending on whether the performance is judged against an internal or external frame of reference (Pieschl, 2008). Viewing calibration through the model of reciprocal determinism centers the critical idea that calibration accuracy results from and influences a multitude of factors.

Calibration of performance is critical to learning. According to Bandura (1986), academic achievement is predicated on the ability to accurately judge one's capability. Generally, research across domains has established the link between calibration and self-regulation. In the specific context of text comprehension, there is evidence supporting that effective metacognitive monitoring and control leads to better comprehension performance (Coiro & Dobbler, 2007; Labuhn et al., 2010; Stone, 2000). For example, Labuhn and colleagues (2010) studied the influence of feedback on calibration accuracy and consequently on the acquisition of mathematics skills with fifth-graders. The students were shown a list of mathematics problems and asked to rate their confidence in solving the problems. They marked their judgment on a 9-point scale ranging from 1 = "definitely not confident" to 9 = "extremely confident." Calibration accuracy was calculated in two steps. First, the researchers computed a bias score by subtracting the performance score from the confidence rating. Second, they determined the magnitude of judgment error by subtracting the absolute bias score from 8. Thus, calibration accuracy scores ranged from 0 (complete inaccuracy) to 8 (complete accuracy). The results from the study showed that feedback improved calibration accuracy, but there was no statistically significant improvement of mathematics skills. However, when Labuhn et al. considered overconfident students, meaning they had overestimated their performance

levels and were poorly calibrated, feedback positively influenced calibration accuracy and they scored significantly higher on the mathematical problem-solving measure.

Schneider and Laurion (1993) investigated the relation between calibration of performance and comprehension in the context of a radio news broadcast. Undergraduate students listened to informational and editorial content that the researchers produced to mimic radio news broadcasts. Schneider and Laurion also manipulated the level of involvement of each recording by creating two versions of the same story. In the high involvement condition, participants were made to believe that the communicated change would impact them. The low involvement condition communicated the same message but was personally irrelevant.

After listening to the audio recordings, the participants took a multiple-choice comprehension measure. As the undergraduates answered each multiple-choice item, they were directed to rate their confidence that their answer was correct on a 7-point Likert scale. The researchers used the Goodman-Cruskal gamma correlation to compute calibration, which indicated the relation between confidence ratings and performance accuracy. The results revealed a strong positive relation between confidence and accuracy, such that high scores were related to high confidence levels and vice versa. This relation held across message formats (informational versus editorial) and involvement conditions (low versus high). Unsurprisingly, calibration was worse for more difficult items, but there was no difference in degree of calibration as a function of individual performance. In other words, undergraduates who performed well on the comprehension measure had the same calibration level as those who did poorly.

Ackerman and Goldsmith (2011) conducted a study to explore differences in metacognitive monitoring on screen versus on paper and its consequences on study behaviors. Undergraduate students read expository texts for the purpose of learning on paper or screen. Next, the participants were directed to predict their performance on a comprehension test comprised of two types of questions—memory of details and higher-order comprehension. Two separate questions were posed: (a) "What percentage of the questions that require memory of details do you expect to answer correctly?" and (b) What percentage of comprehension questions do you expect to answer correctly?" The students indicated their predicted performance on a continuous scale, starting from 25% and ending at 100%. Calibration of performance was the difference between the mean overall performance prediction rating and the test score for each participant. A negative score was indicative of underconfidence, whereas a positive score reflected overconfidence. Although students were overconfident overall, those reading on screen exhibited higher overconfidence than those reading on paper.

Singer and Alexander (2016) uncovered that undergraduates were poorly calibrated in terms of which medium (paper or digital) supports comprehension better. They recruited undergraduate students to read newspaper articles and book excerpts in print and digitally. Students reported their preferred medium for seven different situations that comprised leisure and academic reading contexts. Post-reading, the students completed a comprehension measure and judged the medium in which they performed best. Calibration accuracy was assessed by comparing actual performance with the judgment of performance by medium. Although comprehension performance was better with print, students perceived that they had performed better after reading digitally.

Moreover, contrary to findings from the Ackerman and Goldsmith (2011) study, Singer and Alexander found that students preferred the digital medium. This medium preference was particularly strong when the task was academic.

Given the evidence supporting the importance of calibration of performance accuracy to effective text comprehension, the variability of calibration as a function of medium of text delivery, and its interrelations with other variables, such as self-efficacy, item difficulty, it is imperative to study it in the context of print and audio comprehension. The importance of investigating how this variable plays a role is intensified by the lack of research in the print versus audio medium literature (Clinton-Lisell, 2021; Singh & Alexander, 2022).

Summary

In this section, I reviewed purposefully selected research on key learner characteristics related to cognition (i.e., prior knowledge), motivation (i.e., self-efficacy, reading and listening habits), and metacognition (i.e., calibration of performance) and their influence on comprehension. The opening part of the section briefly covered age of the learners to highlight the importance of the selected variables and their relation to comprehension for undergraduates who are proficient readers and listeners. Notably, as advanced readers and listeners, undergraduates' ability to comprehend complex texts is predicated on a multitude of cognitive, motivational, and metacognitive variables. This contrasts with younger readers and listeners who rely more on fundamental processes such as decoding letters and encoding sounds.

The research has unequivocally established that prior topic knowledge, self-efficacy, reading and listening habits, calibration of performance explain variance in text

comprehension. Nevertheless, it is still unknown if and how these variables work together in different medium presentation conditions to influence comprehension. Therefore, in this dissertation study, I investigated the co-influences of cognitive, motivational, and metacognitive variables on comprehension outcomes across audio and print mediums.

Processing Behaviors, Medium, and Comprehension

Text Processing Behaviors

In the 1980s, reading researchers began to shift focus to *text processing* due to advancements in experimental technologies, such as eye-tracking (Just & Carpenter, 1980; Olson et al., 1983). The fundamental motivation behind this research was to lay bare *what* learners do as they proceed through the text. Unsurprisingly, Kintsch's (1988) construction-integration model of comprehension was developed at this time. The Construction-Integration model centers on meaning-making, a process whereby individuals construct mental representations, which may or may not result in external products. This era marked a pivotal point in the history of comprehension research. It permitted researchers to develop a picture of the individual engaged in the activity of comprehension and the inevitable skills and strategies involved.

Several methodologies have been used to study online comprehension behaviors, and each offers different insights. Eye-tracking, which entails recording eye movements as individuals read texts, revealed a systematic relation between eye fixations and comprehension (Just & Carpenter, 1980). For example, regressive eye movements may be observed when readers encounter semantically related sentences. Put differently, when readers come across two sentences that are related in meaning, they may re-read the previous sentence to interpret or construct meaning. Further, when readers are faced with

less frequent words, they take longer to read them (Just & Carpenter, 1980; Rayner & Duffy, 1986). More recently, fMRI technology has deepened our knowledge of specific brain region networks involved in comprehension processes (Bookheimer, 2002; Keller et al., 2001).

Other commonly used methods to assess text processing have involved think-alouds (Olshavsky, 1976, 1977), interviews, self-report surveys, and trace and log data in the case of digital reading (Kucan & Beck, 1997). While think-alouds have the advantage of giving access to a student's thinking in real-time without memory lapses (Ericsson & Simon, 1980), they can be disruptive, rely on an individual's ability to verbalize their thinking, and do not provide insights into strategies that may be more automatic (Afflerbach & Johnson, 1984; Garner, 1987; Smagorisky, 1989). Interviews, where students respond to questions, are less disruptive, encourage more reflective practices, and permit researchers to gain deeper insights into students' perspectives on their abilities (Garner & Kraus, 1981). Another advantage of interviews is that they can be tailored to a specific individual and allow follow-up probing in cases where the responses are unclear or need clarification (Seda & Pearson, 1991).

Self-report surveys also fall in the category of non-disruptive methods. However, they suffer from the disadvantage that they result in restricted variability in responses as a function of the items (Scott, 2008). Nevertheless, self-report surveys have been used in medium-related research to learn about processing behaviors when researchers are interested in learning about specific processes (Liu, 2005; Murphy et al., 2003). Liu's (2005) investigation of change in reading behaviors in the digital age uncovered that screen-based reading is categorized by increased use of strategies specific to the digital

reading context, such as scanning, keyword spotting, non-linear or hypertext reading, and selective reading.

Text processing behaviors have also been investigated through screen-recordings and capturing log data (Dinsmore et al., 2015; List & Alexander, 2017; Piolat et al., 1997). These data have the dual advantage of being unobtrusive and more trustworthy as they directly record behaviors. For example, List and Bellinger (2019) used a combination of self-reports, log, and trace data to study the difference in strategy as teachers processed information via static text and videos to develop instructional scripts. They collected data on strategies through a strategy use inventory (self-report), time spent on the two information sources (log data), and navigation behaviors captured via screen recording. Example navigation behaviors that the authors coded for were: scanning forward and backward, pausing, changing volume levels. They explicitly stated that log data and screen recording was chosen because of its non-disruptive nature when compared to think-aloud protocols.

Given the advantages of directly recording text processing behaviors by capturing log and trace data and the insights it provides into how learners' interactions with text influence comprehension, I collected data through screen recording on Zoom® as students read and listen to the text. I will record the time taken to read and listen and code for different behaviors. This data were then used to identify different profile groups. After identifying profiles, the screen recording were employed during the interview stage where specific segments were used to cue students to express what they were doing, thinking, and feeling while processing text in print and audio.

Test Processing Behaviors

There has been an interest in test processing or test-taking behaviors since the late 1920s (Lehman, 1928; Lowe & Crawford, 1929). The foundational issues that have animated the research in test processing relate to: (a) improving students' performance; and (b) ensuring the validity of the test scores. The first line of research has concerned itself with uncovering effective test-taking strategies and designing interventions to improve scores. The second strand has investigated behaviors to ensure that the test scores accurately estimate ability and are not capturing other constructs. For the first group of studies, variables such as answer changing have been used as a proxy for uncertainty, engagement, and monitoring (McMorris et al., 1987; Wirtz et al., 2010). In the second group, variables like time taken to complete the test and responses on surveys have been used to gather information on effort, behavioral engagement, and motivation (Arvey et al., 1990; Lee & Jia, 2014).

Studies investigating test processing to study uncertainty draw on decision-making and metacognition research (Bol & Hacker, 2012; Tversky & Kahneman, 1973). The fundamental question has been around answer changing, revising, or reviewing. In early studies, researchers sought an answer to whether or not changing answers leads to better performance (Benjamin et al., 1984; Waddell & Blankenship, 1994). Building on this, an entire edifice of research studies was developed to unpack who, under what conditions, and why students revise answers, and how this may influence performance (Couchman et al., 2016; Lewandowski et al., 2012; Prinsell et al., 1994). A review conducted in the 1980s by Benjamin et al. (1984), synthesized results from 33 studies and found that most revisions are from incorrect to correct and that most students who change their answers improve performance. Benjamin and colleagues noted that the most

frequent method of detecting answer changes had been to observe erasures or crossouts. This review debunked the myth that initial answers are more likely to be correct.

A more recent study conducted by Couchman and colleagues (2016) exemplifies how this line of research has evolved. Couchman et al. also studied the first-instinct fallacy (Kruger et al., 2005; Tversky & Kahneman, 1973), which is the phrase used in the decision-making literature to mark the notion that initial answers have a higher likelihood of being right. Couchman et al. also considered the role played by metacognitive monitoring. Thus, the goal of their study was to determine if metacognitive judgments aid in making revisions that result in the correct answer. In this way, they added an informative dimension to the previous studies.

Specifically, Couchman and colleagues provided insight into one aspect of the mechanism that fuels effective revisions—metacognition. Undergraduate students took a normally-scheduled exam (high-stakes as it impacted grades) and next to each of their response indicated their level of confidence on the correctness of the response (known or guessed). Further, if they made any changes to their initial response, they were instructed to cross out the first response instead of erasing it. In contrast to previous studies, the researchers found that both revising and sticking to the original response was related to correct responses. It was better to stick to responses students were confident about and revise uncertain responses. Metacognitive monitoring was the factor that interacted with revisions (or lack thereof) to influence performance.

The studies that have investigated test processing to study effort have been built on the literature on measurement (Harmes & Wise, 2016; Lee & Jia, 2014; Wise & Gao, 2017). The fundamental question has been around understanding test-taking behaviors

that may infringe upon the measurement of ability. In other words, the purpose of this line of research has been primarily to reduce noise in the performance data. Recently, Rios (2021) meta-analyzed 53 studies containing interventions to improve test-taking effort in low-stakes assessments. Examples of low-stakes, group-based assessments are standardized tests as they do not impact academic grades.

Similarly, experiments conducted by researchers where performance data is gathered through assessments are also low-stakes in nature. The challenge with these types of assessments is that students do not have an incentive to expend effort. Therefore, the results may not be a reliable indicator of their ability. Rios grounded their meta-analysis in expectancy-value theory noting that test effort is influenced by how well the student expects to perform on a test and the value placed on the assessment task (Eccles & Wigfield, 2002; Wise & DeMars, 2005). This meta-analysis synthesized how test-taking effort has been measured and the role of contextual factors like test content, format, and participant attributes in moderating the impact of interventions. They noted several limitations in the research, but two merit highlighting. First, 68% of studies relied on self-report measures of test-taking effort. The most common method in those studies was to use the Student Opinion Scale (SOS) developed by Sundre and Moore (2002). The remaining 32% of studies used response time data. Second, most studies did not account for participant attribute factors in explaining test-taking effort differences.

Wise and Kong (2205) investigated the viability of item response times as a measure for examinee motivation in computer-based tests. The hypothesis driving their study was that unmotivated students would answer questions too quickly (rapid-guessing behavior). The participants for the study were drawn from a population of undergraduates

who had taken the Scholastic Assessment Test, which served as a measure of academic ability. Students took an 80-item information literacy test comprised of multiple-choice questions, and the response time for each item was measured. They also self-reported their effort on the Student Opinion Scale. Using each item's response time distributions and text length, the researchers computed thresholds to differentiate between rapid guessing and solution behaviors. They found that response time scores were a reliable and valid measure of effort. There was a high degree of reliability (alpha = .97). The response time scores correlated with self-reported test-taking effort and person fit (convergent validity) and did not correlate with academic ability (discriminant validity).

This subsection on test processing has been structured around the discussion of two processing indicators—revisions and response time. Given the maturity of the research on test processing, there is a keen sense of the current gaps in the literature. The first set of gaps that this review highlights relate to measurement. In the revision literature, the methods to capture student answer changes have been quite rudimentary and have not changed drastically since the earliest period. Researchers still count erasure marks or ask students to cross out their answers. In the studies capturing effort, there is an overreliance on self-report inventories that limit the insights that can be raked from the data. Evidently, there is a need to employ more reliable methodologies that provide richer information. Therefore, I propose screen recording to gather data on test processing (time, revisions, reviewing, re-framing) and follow-up with cued retrospective interviews to understand the reasons motivating students' behaviors instead of using self-report measures. For example, if a student changes their answer, I will ask them why and

determine if it was because of guessing, misreading, or more laborious cognitive thinking.

The second set of gaps identified in the literature relates to the limited consideration of contextual factors in test processing. The research field is aware of the importance of contextual factors such as item format, test content, test medium on differences in test processing (DeMars, 2000; Dudycha & Carpenter, 1973; Horkay et al., 2006). In other words, attention has been paid to endogenous (test-related) factors. However, there is little consideration of exogenous factors that are not tied to the test context. Specifically, there is no research to the best of my knowledge investigating whether the medium that delivers the content to be learned influences subsequent test processing behaviors.

If effort is indeed conditional on expectancy for success as predicted by expectancy-value theory, then it is likely that the medium of text delivery could indirectly influence test-taking behaviors. In the case of text comprehension, this could happen in cases where the medium influences the degree of comprehension and, as a consequence, the expectancy for success (Wigfield & Cambria, 2010). This is a pertinent area to focus on in the 21st-century context where digital mediums (audio, screen) are becoming more mainstream. Therefore, in this study, I investigated if and how test processing behaviors differ across mediums while considering other learner and text-related factors. By measuring test processing behaviors, I was able to paint a more detailed picture of differences in student behaviors and comprehension across mediums.

Summary

Processing behaviors provide insight into learner-text and learner-test interactions that can deepen our understanding of comprehension processes. Thus, text processing has been extensively studied. In the literature on medium effects, behaviors such as scrolling, reading time, scanning have been investigated in an effort to understand how the affordances of a medium influence comprehension outcomes. These investigations also have implications for informing how the design of text mediums can be optimized to support learning.

The test processing literature has a long history, and the behaviors that have been documented are meant to capture engagement, value for the task, and metacognitive monitoring. Despite the well-established nature of this research area, it has not yet been considered if and how accessing text from one medium or another could influence subsequent test-taking behaviors. Given the documented importance of these behaviors and the apparent lack of research on differences as a consequence of different mediums, in this study, I also recorded test processing behaviors. Both text and test processing behaviors were used to: (a) identify reader and listener profiles, and (b) cue retrospective interviews.

Conclusion

The opening quote is from Walter Kintsch's presentation at the American Education Research Association (AERA) conference. Nearly four decades ago, Kintsch stood in front of the conference attendees in San Francisco, sharing his latest comprehension research using transparencies and a projector. Much has changed since then. The mediums for text delivery have fundamentally altered the quality and quantity

of text we access. However, much remains the same. For instance, our cognitive architecture dictating how and what we learn, our motivations, the nature of writing encapsulated in genres have remained stable. And it speaks to Kintsch's research prowess that his work continues to provide theoretical guidance to inform the design of investigations that expand our understanding of comprehension across mediums of the 21st century.

Indeed, "modeling comprehension in *all* its complexity is impossible or very hard to do," but using the appropriate inquiry methods, one can hope for a reasonable approximation. Therefore, I conducted a mixed methods design where the person-centered quantitative analysis bolstered and was bolstered by the qualitative interviews. Further, following the dictates outlined by Jenkins and Kintsch, I "decomposed" the problem by identifying critical components that feed into the construct of comprehension to shed light on *how*, *why,* and *for whom* it differs across print and audio medium. I explored how learner's cognitive and affective characteristics (e.g., prior topic knowledge, self-efficacy), text processing behaviors (e.g., scrolling, re-listening), and test processing behaviors (e.g., test duration and revisions) result in different profile groups across mediums. This analysis was followed up by investigating how these profiles differ on comprehension outcomes and calibration of performance, and if prior topic knowledge predicts profile membership. Finally, to sharpen our understanding of comprehension across mediums, learners participated in retrospective interviews. This investigation is the first to explore the question of comprehension across print and audio in *at least* some if not "*all its complexity*."

CHAPTER III: METHODOLOGY

Purpose

The primary goals of this dissertation study were twofold: (a) describe the differences in text comprehension across print and audio mediums, and (b) gain a nuanced understanding of how undergraduate students form unique profiles based on their characteristics and processing behaviors in audio and print mediums. Toward this end, I investigated the interrelations among learner characteristics, processing behaviors, and predicted and actual performances on comprehension outcomes when text is processed via print and audio mediums using a mixed method design. I employed finite mixture modeling (FMM), a person-centered statistical technique, as the quantitative tool to classify students into reader and listener profile groups based on their reported self-efficacy, text and test processing behaviors. This analysis is in line with the conceptual model (Figure 1), which shows that comprehension is a multidimensional construct that rests on a complex interplay among learner, text, task, and test features. Further, I qualitatively explained and elaborated text processing across print and audio mediums by analyzing data gathered from cued retrospective interviews with select participants drawn from each of the resulting profiles (van Gog et al., 2005; Salmerónvan et al., 2017).

Design

Given the focus on exploring and understanding differences and similarities in text comprehension across print and audio mediums, this dissertation study employed a sequential explanatory mixed methods design, which is visually depicted in Figure 2 (Creswell, 2009). The goal of this design was *complementarity,* wherein the qualitative data were used to enhance the understanding of text comprehension across print and

audio, and to elaborate on the profile groups formed from the quantitative data for each medium (Greene et al., 1989). As shown in Figure 2, there were two points in this mixed methods study where the quantitative and qualitative phases connected and complemented each other. First, the sample for the qualitative phase was selected based on the results from the quantitative analysis. Second, the results from the quantitative and qualitative phases were combined to elaborate our understanding of text comprehension across mediums, and the reader and listener profiles.

Each student processed three comparable text segments from the same source (Pinker, 2018); one segment per condition—reading, listening, and reading aloud—leading to a within-participant design. Reading and listening were the experimental conditions, and the reading-aloud condition was only be used to collect baseline data on reading time. The decision to include a reading-aloud condition in the study was based on observations made during a pilot study. Specifically, preliminary results indicated that many students in the pilot study were skimming or skipping the text, as evidenced by their relatively short reading duration. To address this issue and obtain a more accurate measure of reading duration, a reading-aloud condition was included in the current study.

It is important to note that the reading-aloud condition was not intended to precisely measure reading duration in all contexts. Rather, it was included to provide a rough approximation of the amount of time students take to read the text if they read every word. It is recognized that reading aloud and reading silently may lead to different reading times. Nonetheless, the reading-aloud condition was included to provide a more comprehensive understanding of students' reading behavior in the context of the study.

The text segments and conditions were counterbalanced using a Latin square design (see Table 1) to protect against order effects (Grant, 1948).

Figure 2

Visual Representation of the Sequential Explanatory Mixed Methods Design

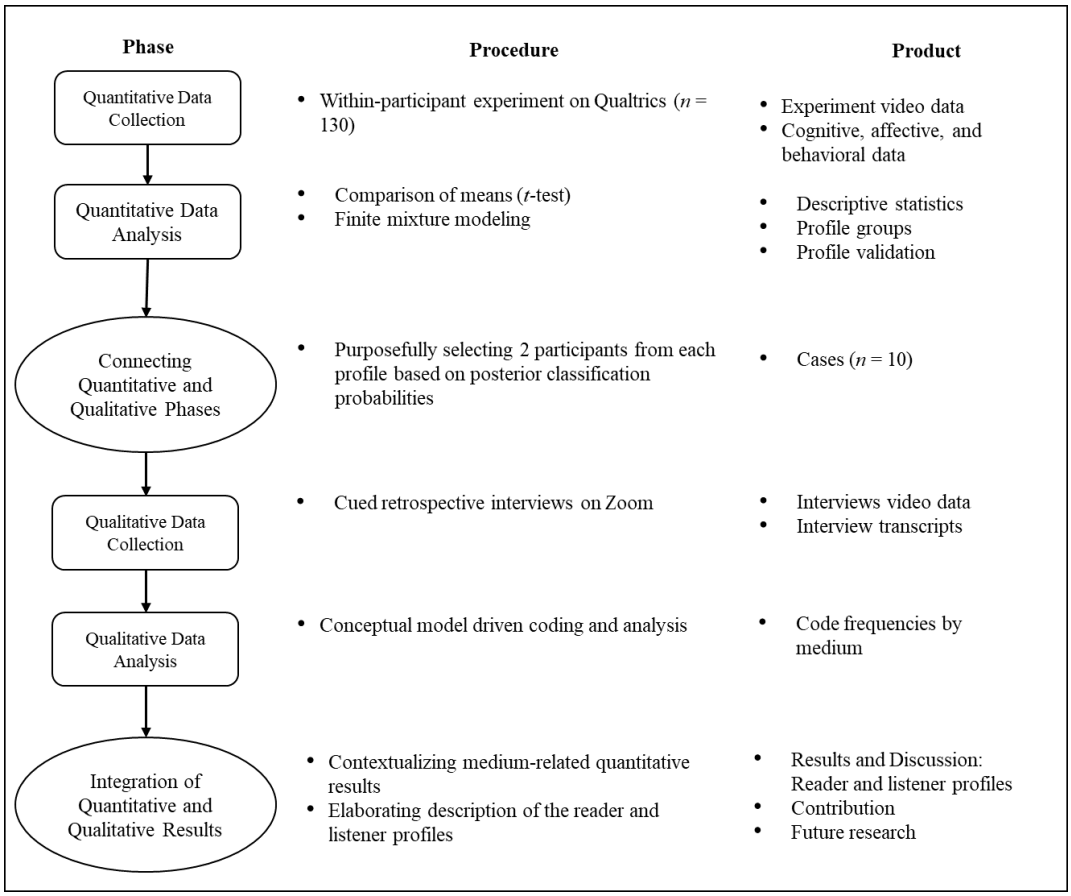


Table 1

Latin Square Design Depicting Counterbalancing of Conditions and Text Segments

Text Segments	Conditions		
	Reading Aloud	Reading	Listening
Inequality	A	B	C
Happiness 1	B	C	A
Happiness 2	C	A	B

The quantitative portion of the study included data on variables related to the learner, task, and test. The variables related to the learner included: (a) self-efficacy in

reading and listening; (b) reading and listening habits; (c) demonstrated topic knowledge. Task variables consisted of: (a) mediums (audio and print); (b) duration (reading and listening); (c) text processing behaviors (e.g., scrolling, rewinding); and (d) self-reported task experience (e.g., task enjoyment and difficulty). Variables under the test category included: (a) test-taking times; (b) test processing behaviors (e.g., revising, pausing); (c) scores for each level of comprehension—main or gist understanding, recall of explicitly stated details or low inference of details, deeper level inferencing, and vocabulary gain; and (d) calibration of performance.

The qualitative data were collected through cued retrospective interviews (van Gog et al., 2005; Salmerónvan et al., 2017). Participants were randomly selected from each of the profile groups identified using finite mixture modeling and invited for interviews. Participants who agreed to be interviewed were shown specific parts of their video recordings corresponding to purposefully selected text segments from both the reading and listening conditions. Immediately after viewing each video clip, participants were asked to verbalize what they were doing, thinking, and feeling.

Method: Quantitative Phase

Participants

A total of 130 undergraduate students enrolled in human development and educational psychology courses participated in this study. Three students did not record their videos and were excluded from the finite mixture modeling. In the sample, 109 students identified as women, 18 as men, and 3 preferred to self-describe. The overall mean age was 20.4 (SD = 1.87). Of the undergraduate sample, 15 were first-year students, 42 were sophomores, 36 were juniors, and 37 were seniors. The sample

represented a wide range of majors—Arabic, Architecture, Biology, Business Management, Community Health, Computer Science, Criminology, Economics, Education, English, Family Science, Finance, Government and Politics, Hearing and Speech Sciences, Human Development, Information Sciences, International Relations, Jewish Studies, Kinesiology, Mechanical Engineering, Nursing, Music Education, Physiology and Neurobiology, Psychology, Public Health Science, Sociology, and Studio Art.

Majority of the students reported English as their native language (n = 109). Other native languages reported in this study were: Chichewa (n = 1), Chinese (n = 5), Gujarati (n = 1), Hindi (n = 1), Italian (n = 2), Korean (n = 3), Nepali (n = 1), Spanish (n = 4), Telugu (n = 1), Turkish (n = 1), and Vietnamese (n = 1). Undergraduate students were chosen as the population of interest because they have demonstrated the ability to process complex texts dealing with abstract concepts under multiple mediums (Singer Trakhman, Alexander, & Berkowitz, 2017c; Singer Trakhman, Alexander, & Silverman, 2018). With mature readers, I was able to go beyond foundational linguistic skills such as decoding and fluency (Fuchs et al., 2001; Gough & Tunmer, 1986; Perfetti, 1985) and examine the intertwined influence of learner and task-related factors such as interest, self-efficacy, and processing behaviors on comprehension performance (Alexander, 2005).

The participants were recruited primarily from undergraduate courses offered by the Department of Human Development and Quantitative Methodology at the University of Maryland. They were given extra credit for participating in the quantitative part of the study and entered into a raffle to win $100 Amazon gift card for participating in the

interviews. The study was reviewed and approved by the university's Institutional Review Board.

Experimental Texts

Segments from *Enlightenment Now: The Case for Reason, Science, Humanism, and Progress* (Pinker, 2018) were used as the experimental texts (see Appendix A). This book is a mixed genre text (Alexander & Jetton, 2000). It contains factual information typical of exposition but is written with narrative elements such as dialogue and the narrator's viewpoint with the purpose to inform and entertain.

Three segments were drawn from two book chapters entitled "Inequality" and "Happiness." Segment 1 was from Inequality, and Segments 2 and 3 were taken from Happiness. Each segment is complete in itself and does not require knowledge of the content of other segments. These text segments were matched on length of text and duration of audio, readability levels, the number of paragraphs, and words per minute (see Table 2).

The audio versions of the experimental texts were taken from the audiobook narrated by Arthur Morey, an award-winning male narrator of over 400 audiobooks from the United States (AudioFile, 2023).

Table 2

Properties of the Text Segments Included in the Study

Text Segment (No.)	Word Count	Paragraph Count	Audio Duration	Words Per Minute	Flesch-Kincaid Readability
Inequality (1)	1182	7	8m7s	145.6	14.2
Happiness (2)	1237	8	8m7s	152.4	11.4
Happiness (3)	1197	8	7m57s	150.5	12.3

Further, in accordance with Latin square design, approximately one-third of students first read Segment 1 aloud, then listen to Segment 2, and finally read Segment 3 silently. The next group of participants first listened to Segment 3, then read Segment 1, and finally, read aloud Segment 2. The remaining one-third of the participants first read Segment 2, then read aloud Segment 3, and finally, listened to Segment 1.

Pre Reading and Listening Measures

Demographics Questionnaire

Participants completed a demographic questionnaire pertaining to their personal and educational background (see Appendix B). In this questionnaire, participants reported their gender identification, age, native language, class standing, and disciplinary major. This part of the study was completed on Qualtrics®.

Self-Efficacy for Reading and Listening

The self-efficacy measure was adapted from Shell et al. (1989). The original reading self-efficacy measure was shortened in terms of number of items for the purpose of not overtaxing the participants. The original reading task scale was modified to tap into self-efficacy in both reading and listening. As shown in Figure 3, participants rated their self-efficacy on a 100 mm line. The maximum self-efficacy score was 800. Each student's score was scaled between 0–1 by dividing the actual score by the maximum possible score.

The reliability of the original scale, calculated in terms of Cronbach's alpha was 0.91. Internal consistency reliabilities for the adapted reading and listening self-efficacy scales used in this study in terms of Cronbach's alpha were 0.87 and 0.90, respectively.

Two sample items from the eight-item scale are shown in Figure 3 with the full scale included in Appendix B.

Reading and Listening Habits

Participants reported their reading and listening habits by providing information related to hours spent reading and listening, and quantity of reading/listening material in possession. Sample items for the Reading and Listening Habits are shown in Figure 4.

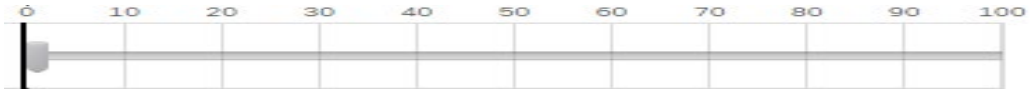
Figure 3

Sample Items from the Reading and Listening Self-Efficacy Measure

Direction. Rate your confidence in being able to read/listen and understand what the author/narrator was saying in each of the following texts: (0 – not confident at all; 100 – highly confident)

a) A short fiction story.

- *Print*



- *Audio*

b) A textbook in your major field.

- *Print*

- *Audio*

Figure 4

Sample Items from Reading and Listening Habits Measure

In a typical week, how many hours do you spend reading the following:

Reading for leisure

- *Books (e.g., novels, non-fiction, biographies)*_____
- *Other forms of leisure reading (e.g., news stories, magazine articles, op-eds, blog posts)* _____

Reading for school

- *Textbooks*
- *Other forms of school reading (e.g., teacher handouts, slides, assigned readings)* _____

In a typical week, how many hours do you spend listening:

Listening for leisure

- *Audiobooks (e.g., novels, non-fiction, biographies)*_____
- *Podcasts* _____

Listening for school

- *Audio textbooks*
- *Other forms of school listening (e.g., using text to speech feature for slides, teacher handouts)* _____

Approximately how many of the following forms of text, do you have in your possession?

Books: ☐

e-books: ☐

Audiobooks: ☐

Topic Knowledge Test

In order to determine the students' topic knowledge, a specialized vocabulary test was administered. The test contained six vocabulary items, two from each of the three segments, that the students were directed to define to the best of their abilities. The words in the topic knowledge test were *paradox, materialism, warped, masochistic, fallacy, and egalitarian.*

To create this prior knowledge test, I followed a series of steps. First, a list of 12 potential words (four from each segment) that were deemed critical to understanding the meaning of the text segments was compiled in consultation with a panel of experts. Next, this 12-item test was piloted with nine undergraduate students. Finally, to compile the shortened list of words, I retained only items with three or less missing responses and items that showed variability in the degrees of correctness of the responses.

Participant responses on the final measure used in the study were scored on a scale of 0–2. Full score of 2 was awarded to those students who approximated the word meaning. A score of 1 indicated some but not full understanding of the meaning of the word. A score of 0 was given if the student left the question unanswered or provided a definition that was entirely incorrect. The range of possible scores was 0–12. The final score was scaled to a 0–1 range by dividing the assigned score by the maximum score. This was done to ensure that scores were on the same scale as the measures in this study.

Interrater agreement was 98.7% for 20% of the scoring. Here I have displayed two items from the Topic Knowledge measure in Figure 5 and the full measure provided in Appendix B.

Figure 5

Sample Items from Topic Knowledge Measure

Topic Knowledge Items

Directions. Provide a definition for each of the words displayed. If you are not certain of the definition, share what you think the meaning of the word may be.

Q1. paradox

Q2. egalitarian

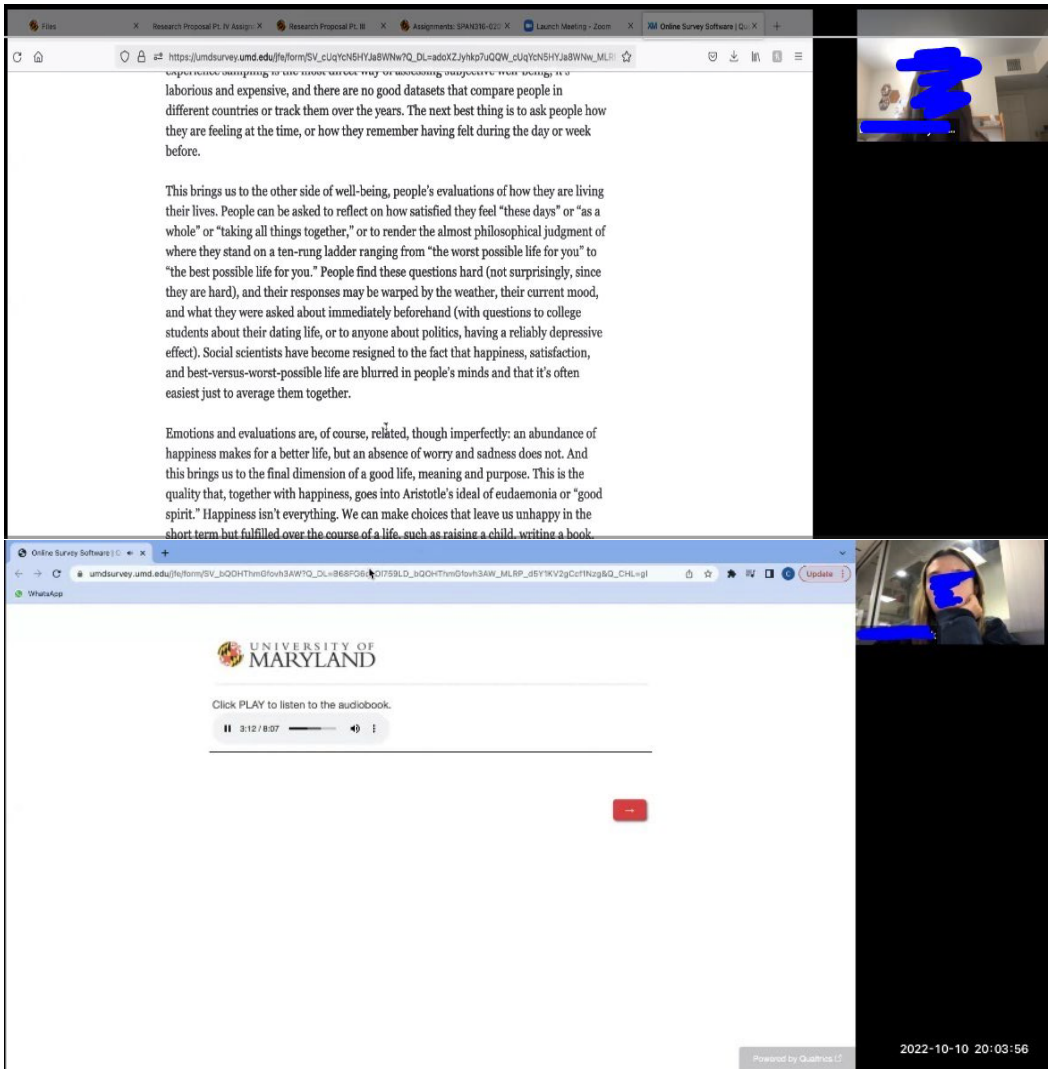
During Reading and Listening Measures

Text Processing Behaviors

Students completed the reading and listening phases in Qualtrics®. As shown in Figure 6, both the print and audio text were presented on the Qualtrics® platform. Students shared and recorded their screens and videos of themselves as they completed the tasks using Zoom®. I captured reading and listening durations through Qualtrics®. I coded the videos at 3 sec intervals for frequency of other types of task behaviors using Atlas.ti® (Bakeman & Gottman, 1997): (a) forward and backward scrolls; (b) following the text with the cursor (cursor reading); (c) using the cursor to highlight text; (d) keeping the screen stationary; (e) off-task behaviors (e.g., eating); (f) gazing patterns; (g) increasing playback speed; (h) fidgeting; (i) forwarding and rewinding audio. Interrater agreement was established with 15% of the data and found to be 94.7%.

Figure 6

Students Completing Reading and Listening Tasks on Qualtrics and Recording Screen on Zoom®



Post Reading and Listening Measures

Comprehension Test

Comprehension measures were administered immediately after reading and listening to the texts. The test includes both selected-response and constructed-response items. The selected-response items were in multiple choice format. They targeted recall of information that was either explicitly stated in the texts or required low level

inferencing. An additional set of MCQs tapped into higher order inferencing. The constructed-response items were in short answer format. All items were presented in digitally on Qualtrics®. The complete list of items is given in Appendix C.

Selected-Response Items. There were seven selected-response items associated with each text. Four of these items tapped into NAEP's cognitive target *recall* (National Assessment Governing Board, 2015) and accordingly required participants to recall explicitly stated information or make a low-level inference. The answer choices for each item were on a graduated response scale (Alexander et al., 1998) representing four categories based on text understanding (good or poor) and response (correct or incorrect). The first category is good understanding of the text and correct response, which was scored as 4. The second category is good understanding of the text but incorrect response and was awarded a 2. The third category, which was given a score of 1, represents poor text understanding and correct response resulting from the faulty understanding. The final category is poor text understanding and incorrect response, which resulted in a 0 score. The range of possible scores for this section was 0–16. The final scores were scaled from 0–1 by dividing actual performance by maximum possible score.

The remaining three multiple choice questions for each text segment assessed inference making, part of the *interpretation* cognitive target identified in NAEP's *Reading Framework*. The answer choices for each item were on a graduated response scale (Alexander et al., 1998), similar to the ones described above for the recall of explicit information items, representing four categories based on text understanding (good or poor) and response (correct or incorrect). The range of possible scores for this

section was 0–12, which were scaled down to 0–1. Items representing the two cognitive targets are displayed in Figure 6.

Figure 7

Sample Comprehension Questions for the Recall and Inference Cognitive Targets

Recall of Explicit Information Item

Direction. Select the correct response from the options given below.

Named after an economist, the Easterlin paradox states that:

- higher income does not make people happier (Key – correct)
- *economic growth is the primary driver of well-being (Distractor)*
- *countries do not pursue happiness as the goal of development (Distractor)*
- *people in poorer countries are happier than people in richer countries (Distractor)*

Inference Item

Directions. Select the correct response from the options given below.

The author juxtaposes the views of scientists on inequality to:

- *illustrate the limitations in the current research on inequality (Key – correct)*
- *indicate that inequality must be studied alongside other factors (Distractor)*
- *demonstrate how particular social ills arise from inequality (Distractor)*
- *show how scientists can reach different interpretations from the same data (Distractor)*

To develop the selected-response items, I first made sure that the items were assessing the appropriate cognitive target, meaning that the items assessing recall could be answered with information explicitly stated in the text segments. In contrast, the items targeting inference required students to deduce information. Second, the developed items were based on information the students encountered in the text's beginning, middle, and end. The items were tested with a panel of graduate students who read the text segments and provided answers to the measures. This was followed by a detailed discussion of

each item's clarity of language, difficulty level, and cognitive target. The items and response options were revised based on the experts' feedback.

Constructed Response Items. These items align with NAEP's *interpret* cognitive target and will require participants to provide short answers. The items tested gist or main idea understanding and ability to infer meaning from context. Gist understanding was assessed through asking students to identify the main idea of each passage (see Figure 7). To ensure that the participants understood the meaning of "main idea," a brief definition was provided as part of the question prompt. The main idea for the segments was determined a priori by me and several graduate students and faculty with expertise in text structure. Students typed in their responses on Qualtrics®.

The responses were scored on a scale of 0–2, which were scaled down to 0–1 range to maintain consistency across all measures in the study. Full score of 2 was awarded to responses that captured the essence of the main idea for each text segment and were a complete answer, that is they do not require the assessor to fill in any gaps. An example full score response for the *Inequality* text segment is: *"There are many wrong assumptions about the causes and consequences of inequality. Inequality is not caused by an increase in wealth and it does not lead to unhappiness. In fact, increasing wealth leads to higher well-being."* A score of 1 indicated some but not full understanding of the main idea or a response that is incomplete. An example response in this score category is: "*There are many wrong assumptions about the nature of inequality."* This is a correct statement but does not explicate the nature or type of wrong assumptions made about the nature of inequality. A score of 0 was given if the student left the question unanswered or provided a main idea that was entirely incorrect. For

example, a response such as: *"Inequality leads to unhappiness,"* indicated that the participant had entirely misunderstood the main idea conveyed by the text segment. Following training, two individuals independently scored 20% of the main idea question. Interrater agreement was 92.3%. All differences were resolved in conference.

The vocabulary gain measures each included two of the six words that made up the prior topic knowledge measure namely: *paradox, materialism, warped, masochistic, fallacy, and egalitarian.* Only those words were presented in the vocabulary measure that were encountered in the text read or listened. The initial scoring of the responses was carried out in the same manner as that described for the topic knowledge test. An additional step here was to calculate the vocabulary gain score, that is the change from pre to posttest. A score of 0 implied no change and a score of 4 highest possible gain.

It was expected that after reading a word in context, a student might be able to infer the meaning of the word. For example, the meaning of the word "egalitarian" is inferable from the text. The first clue comes in the form of examples when the text talks about "economically egalitarian countries like France and Sweden." The next clue in when the egalitarian countries are compared to "lopsided" ones like "Brazil and South Africa." A student adept at inferencing could either use prior knowledge about the countries' economic statuses or meaning of the word lopsided to figure out what is meant by egalitarian.

Test Processing Behaviors

The test taking portion of the videos were coded for behaviors that indicated students' test taking approaches. Similar to text processing behaviors, test taking duration data were extracted from Qualtrics®. Instances of other predetermined test behaviors

were coded in the video recordings using Atlas.ti®. Specifically, I coded for the following text processing behaviors in the videos: (a) frequency of revisions (e.g., changing answers in the MCQs), and (b) reframing constructed response items set. Interrater agreement between the two coders from coding 20% of the data was 98.7%.

Figure 8

Comprehension Question for Assessing Gist Understanding

Main Idea Item
What was the main idea of this passage? That is, what was the author trying to communicate? Justify your response.
_____
_____
_____

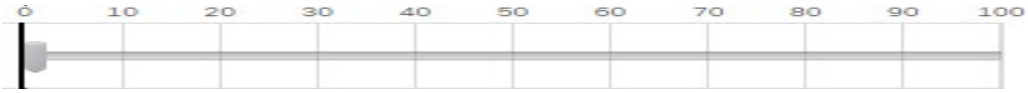
Calibration of Performance

To assess students' ability to judge their performance on the comprehension measures, they were asked to complete a performance judgment item immediately after answering the comprehension questions after the print and audio conditions. The score for this item was used to determine calibration accuracy, which denotes the difference between the actual and perceived performance. Smaller calibration scores correspond to higher calibration accuracy. A score of zero denotes perfect calibration, negative scores are a sign of underconfidence, while positive scores imply overconfidence in performance. The judgment of the performance item is shown in Figure 8.

Figure 9

Item for Assessing Calibration of Performance

Direction. On a scale of 0 to 100, rate your performance on the comprehension questions you just answered (0 –very poor; 100 – excellent):



Post-Task Completion Measures

Self-Reported Task Experience

For this measure, students were asked to rate the enjoyment and difficulty of the task for each medium (see Appendix D). Items for this measure are shown in Figure 9.

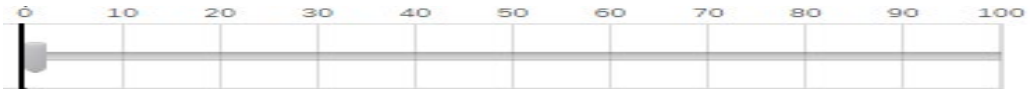
Figure 10

Items from the Task Experience Measure

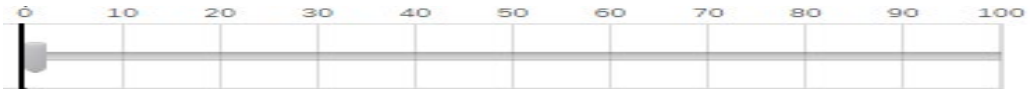
Direction. On a scale of 0 to 100, rate how enjoyable and how difficult you found this experience:

ENJOYMENT of the experience (0 – not enjoyable at all; 100 – highly enjoyable)

In Print

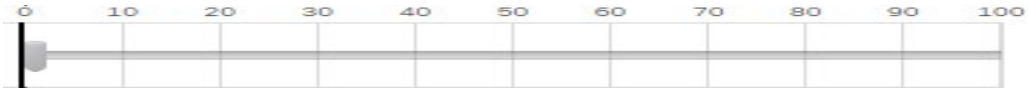


In Audio

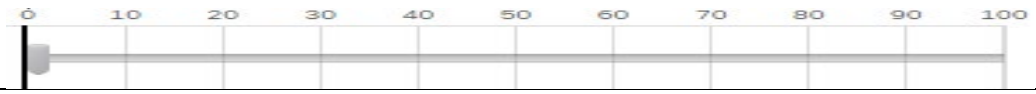


DIFFICULTY of the experience (0 – very easy; 100 – highly difficult)

In Print



In Audio



Procedure

Undergraduate students were recruited from the University of Maryland and offered extra credit for participation. The entire quantitative phase of the study was conducted on Qualtrics®, and the participants shared and recorded their screens using Zoom® while completing the experiment. The session lasted approximately 30 to 40 minutes. The consent form associated with the study was also presented digitally on Qualtrics®. The study unfolded in several stages: pre-text processing, during text processing, post text processing.

In the pre-text processing stage, students completed measures related to demographics, self-efficacy, and reading and listening habits. Then each student read silently, read aloud, and listened to comparable segments from the book *Enlightenment Now* by Steven Pinker in counterbalanced order. They shared and recorded their screens using Zoom®. After reading and listening, participants completed measures related to comprehension performance and judgment of performance while their test-taking behaviors were recorded. Post-task completion participants self-reported their reading and listening experience.

Data Analysis

First, I conducted variable-centered descriptive analyses to uncover the difference across the print and audio mediums (RQ 1). Second, the variables related to the learner, task, and test were used as indicators in finite mixture modeling (FMM), a person-centered statistical technique (RQ 2). The assumption and purpose of FMM align with the exploratory lens adopted in this investigation. FMM assumes that the population contains a mixture of unobserved groups. The goal of FMM is to unearth the number and nature of

the unobserved groups, called classes or profiles (Vermunt & Magidson, 2002). In this study, I use the term profiles to talk about the groups uncovered by FMM.

FMM was carried out separately for the print and audio conditions to identify reader and listener profiles. I fit several models for readers and listeners, extracting one to five profiles. For each of these models, I compared models with four different covariance structures (Bauer & Curran, 2004; Pastor et al., 2007). Model A represented the latent profile analysis model, where indicator variances were constrained to be equal across classes, and covariances among the indicators were fixed at zero. In Model B, indicator variances were estimated freely across profiles, while covariances remained fixed at zero value. In Model C, indicator variances were estimated freely across profiles, and covariances were constrained to equality across profiles. In the final model, Model D, indicator variances and covariances were both estimated freely. FMM was conducted in *MPlus© 8.8* (Muthén & Muthén, 1998-2017).

The estimator for the models was maximum likelihood, which is the default in *MPlus©*. *MPlus©* automatically produced different sets of starting values and reported the solution with the best likelihood. The number of initial stage random starts was set to 20, the number of final stage optimizations was 4, and the number of initial stage iterations was set to 10. To account for non-normality in certain indicator variables (e.g., increasing audio speed, number of audiobooks in possession), these variables were specified as count variables in the model and its estimation. This was done to ensure that the model accurately captured the distribution of these variables and their relation to the reader and listener profiles.

In contrast to more standard cluster analytic approaches, FMM is model-based. Therefore, several fit statistics can be used to select the best-fitting model less arbitrarily. However, it should be noted that model selection remains an open research topic, and that there is no consensus as to which fit index should be used across different conditions (Nylund et al., 2007; Tien et al., 2013).

To statistically identify the number of profiles, I used two fit indices, namely, Bayesian Information Criterion (BIC) and Bootstrap Likelihood Ratio Test (BLRT) (Schwarz, 1978; Tein et al., 2013). BIC indicates the relative fit of the model in comparison to other models. Lower BIC values are indicative of better fit. The BLRT gives relative fit of a model with n profiles in comparison to a model with n-1 profiles. Other sources of information used in model selection were the average class assignment probabilities and theoretical interpretability of the profiles (Pastor et al., 2007). The average class assignment probabilities for all individuals within a profile give the reliability of classifying students into that profile (Geiser et al., 2006). Reliability values at or above .90 are considered excellent (Kline, 2012).

The generalizability of the profiles was evaluated by examining the relation between the profiles and external variables (Pastor & Erbacher, 2019). Scores on the comprehension measures were used as outcome and prior topic knowledge as a covariate to predict profile membership. In FMM, the external variables are included in the model in contrast to cluster analysis, where the relations are tested post hoc using ANOVA type analyses.

Another critical difference between cluster analysis and FMM techniques is that there is a probability associated with a student belonging to a profile in FMM. In contrast,

in cluster analysis, profile membership is deterministic. Consequently, when external variables are included in FMM, it accounts for the uncertainty in student profile classification (Pastor et al., 2007; Vermunt & Magidson, 2002).

Method: Qualitative Phase

Participants

A total of 10 students took part in cued retrospective interviews; seven identified as women, and three identified as men. Eight participants indicated that English was their native language, and two reported Chinese and Korean as their native languages. The class standing of the interviewees ranged from first-year students to seniors. Each participant was majoring in a different discipline—Arabic and Government and Politics, Biology, Computer Science, Economics, English, Hearing and Speech Sciences, Microbiology, Nursing, Physiology and Neurobiology, and Psychology.

Students from each identified reader and listener profiles were invited for interviews through emails. They were offered a chance to win a $100 Amazon gift card for participation. The interviews were conducted and recorded using Zoom® and they lasted approximately 30 minutes.

Measures

Cued Retrospective Interviews. Students watched video clips of their processing behaviors while reading and listening to specific portions of the text. The video clips were pre-selected to play or show certain parts of the text segment. The text segments were selected based on text difficulty due to change in tone, complexity of ideas and arguments. The selected clips are given in Appendix E. The length of the print video clips varied by participant. Each clip was about a minute and a half for audio. After being cued

by the video clip, the participants were asked to share what they were doing, feeling, or thinking while doing the task.

Procedure

The interview sessions were conducted and recorded on Zoom®. Two video clips corresponding to the print and audio conditions from the participant's task recording were pre-selected to cue the participant. Before playing each of the video clips, I provided the following instructions:

I will be showing you short videos of you completing the task. I will pause after playing a segment and ask you to explain what you were doing, thinking, or feeling at that point. Verbalize everything that comes to mind.

If the participants stopped talking for 5s, I prompted them by saying, "*Is there anything else you would like to add about what you were doing?"* before moving to the next video segment. The full interview protocol is provided in Appendix E.

Data Analysis

I followed a *deductive/inductive hybrid analysis approach* (Elo & Kyngäs, 2008; Proudfoot, 2022; Saldaña, 2021). Specifically, I began with four codes that corresponded to the conceptual framework that formed the backbone of this study—*Learner, Task, Text,* and *Test* (Jenkins, 1979, 1986). In the first coding cycle, I ascribed one of the four conceptual codes to each idea unit. Idea unit was defined as an utterance that conveyed a single idea whether in a phrase or a complete sentence. In the second cycle, each idea unit was further sub-coded based on the key variables listed under each of the vertices of the conceptual model *and* what emerged in the interview data (e.g., habits and preferences,

metacognition). Code descriptions with examples of verbalizations drawn from the student interviews are provided in Tables 3a–3d.

Once the data were coded, I analyzed and synthesized them to identify similarities and differences between different groups and the dimensions of the data. To facilitate this process, content analysis was conducted on the coded data, which involved identifying and analyzing patterns and themes that emerged from the data. The data were analyzed with a focus on understanding the similarities and differences between the print and audio mediums, as well as among the three reader profiles and three listener profiles that were identified in the quantitative strand of the study. Through this process of analysis and synthesis, commonalities and differences were identified across the various dimensions and sub-codes. These commonalities and differences were used to generate insights into the behaviors and attitudes of readers and listeners, and to provide a more nuanced and comprehensive characterization of the profiles that were identified in the study.

After conducting two cycles of coding, data were synthesized to qualitatively describe and compare medium differences and the reader and listener profiles. The data from the interviews were transcribed using otter.ai, a built-in application in Zoom® that converts speech to text. The coding cycles were conducted on Atlas.ti®. Along with a research colleague, we coded 20% of the data together to establish interrater reliability. The interrater reliability was found to be kappa = 0.85. All disagreements were resolved in conference.

Table 3a

Learner Dimension

Subcodes	Description	Example Quote
Affect	Verbalization related to general emotional state without referring to the text or the task.	"I have anxiety." "It's also like just a lack of patience at that point of the day"
Habits and preferences	Typical behaviors associated with reading and listening, including but not limited to time spent reading and listening to text. Likes and dislikes related to reading and listening such as preference for a specific medium or genre of text.	"I don't like listening to talk radio or like listening to like someone just like talk at my ear."
Metacognition	Awareness of oneself as a learner, the factors that may impact performance, or the regulation of cognition (Garner & Alexander, 1989)	"I was like trying to gauge how much attention should I pay pay to it." "I'm not a very like philosophical person."
Self-Efficacy	Participant's belief (or lack thereof) in their abilities specifically in the domain of reading or listening (Bandura, 1997).	"I was always able to like read, read the material. so I never chose to do the audio."
Topic/Prior knowledge	Verbalizations referring to pre-existing knowledge or understanding of the text's subject (or lack thereof).	"My major is not relevant to this field."

Table 3b

Text Dimension

Subcodes	Description	Quotes
Affect	Expressions of emotions induced by the nature of the text. The verbalization includes reference to the text.	"Find it [text] like fascination."
Structure	Verbalizations related to the text organization such as how the information is arranged.	"They were connected very like by thin topics."

Themes, Ideas	Expressed themes, ideas, arguments, topics in the text segment	"They mentioned um a third world country where a woman was living poorly."
Tone	Verbalizations associated with the tone of the text, e.g., intended audience, mood, style evoked by the text segment.	"I think they wrote this for scholars."
Syntax	Verbalizations related to the sentence structure encountered in the text segments.	"like this sentence is too big."
Vocabulary	Verbalizations about the words found in the text segments.	"It still had the same jargon of like long word."

Table 3c

Task Dimension

Subcodes	Description	Quotes
Elements, Features	Referring to the affordances of the mediums, or medium-specific features.	"I thought about how the voice seem more robotic, less human like."
Cognitive processing behaviors	Cognitive actions taken to comprehend the text segments, described in the context of the task.	"I was like again trying to find a theme"
Physical processing behaviors	Physical actions taken to assist in making meaning of the text.	"I was multitasking and writing down things"
Task experience	Verbalizations about how the task felt.	"It was kind of boring like sitting there listening"

Table 3d

Test Dimension

Subcode	Description	Quotes
General	Evaluative statement about the general quality of understanding of the text.	"I read it with my eyes, but not my brain."

Specific	Evaluative statement about a particular part of the text.	"And I'm like, I know that." "But I don't know it like I know how to define it."

CHAPTER IV: RESULTS AND DISCUSSION

To answer the research questions posed in this mixed methods investigation about text comprehension across print and audio, I compared variables related to the learner, task, and test across the two mediums (RQ 1). I used finite mixture modeling to identify reader and listener profiles (RQ 2a) and validated the profiles on external variables of comprehension performance and prior topic knowledge (RQ 2b and RQ 2c). Differences between the two mediums and the defining characteristics were further explicated from data gathered from cued retrospective interviews with representative students from the resulting profiles. In addition, those interviews afforded a more detailed description of the quantitatively unearthed reader and listener profiles (RQ 3). I conducted content analysis guided by the four dimensions of the tetrahedral model—learner, text, task, and test.

The results from all these analyses are detailed in this chapter. Specifically, I first present the quantitative results, which include variable-centered descriptive and inferential statistics of the key variables for each medium that informed this investigation. Next, I present the person-centered finite mixture modeling results that led to the identification of reader and listener profiles. The findings from the qualitative analysis of interview data then follow. Finally, I conclude the chapter by integrating the quantitative and qualitative results as dictated by the mixed-methods approach that informed this study.

Quantitative Results

Research Question 1: Variable-Centered Results

Before fitting finite mixture models, variable-centered descriptive analysis was undertaken. The descriptive results are presented in four subsections that follow the

chronology of the experimental task, which unfolded in four parts—*pre*-reading and listening, *during* reading and listening, *post* reading and listening, and *post* task.

Pre Reading and Listening

Notably, students reported higher self-efficacy for reading than listening (Table 4). This corresponded to their relative exposure to print and audio as captured by reported hours spent per week reading and listening. While students read for an average of 14.49 hours a week, they only listened for 5.30 hours, including audiobooks and podcasts. Keeping with this trend of greater experience reading than listening, participants also reported possessing approximately 30 books and eBooks but only 1 audiobook on average. It appears that although audiobook and podcast usage is on an upward trajectory, its adoption still lags far behind print reading for undergraduates (Edison Research & Triton Digital, 2021; Pew Research Center, 2019; Singh & Alexander, 2023). On topic knowledge, as measured by a specialized vocabulary test, students performed slightly above 50% on average.

Table 4

Means, Standard Deviations, and Paired Sample T-Test for Pre-Reading and Listening Measures

Variable	Print		Audio		*t*	*p*
	M	*SD*	*M*	*SD*		
Self-Efficacy	0.81	0.12	0.74	0.16	6.8	<.001
Hours/week	14.49	9.14	5.30	6.76	10.57	<.001
Possess books	22.98 (38.27)					—
Possess eBooks	6.61 (15.15)					—
Possess audiobooks	1.02 (2.98)					—
Topic knowledge	0.58 (0.23)					—

Note. Standard deviation in parenthesis.

During Reading and Listening

Students' reading and listening behaviors were coded from video recordings and are presented in Table 5. Participants spent lesser time reading than listening, with greater dispersion around the mean for reading duration. This was expected given that print affords greater control over speed than audio, which is a more passive act. The only way to speed up listening is to increase playback speed or to skip the audio altogether, which few participants chose to do in this study.

I also coded off-task behaviors while students read and listened. Overall, students engaged in higher off-task activities while listening. They ate or drank, yawned, or did more body movements when listening to audio than reading print. These off-tasks behaviors might be indications that participants found listening more boring (Danckert et al., 2018). The higher incidence of off-task behaviors could also be a manifestation of lack of vigilant attention, a term used to define prolonged attention allocation to unchallenging or uninteresting activities (Langner & Eickhoff, 2013). The qualitative analysis provided more insights into this aspect and is discussed later.

Other text processing behaviors were unique to the print and audio conditions. For example, for print text processing behaviors included scrolling, cursor reading, highlighting, and keeping the screen stationary while reading. Students tended to engage in substantially more forward scrolls (25.67) than backward scrolls (1). The high number of forward scrolls might indicate skim reading patterns associated with screen reading, while the low number of backward scrolls points to the fact that students did not engage in much re-reading across different sections of the text (Liu, 2005; Wolf, 2018). In this study, I only captured scrolling frequency, but I observed that participants had different

scrolling patterns. Some students continually scrolled while reading, whereas others would scroll, pause and read while keeping the screen stationary.

Interestingly, I observed two types of cursor movements during reading, going over each word with the cursor horizontally or following each sentence in a vertical cursor movement. Cursor movement appears to be the digital equivalent of fingerpoint-reading, which is usually seen in beginner readers (Ehri & Sweet, 1991). Given that this study involved undergraduate students assumed to be advanced readers, I did not expect to find this text-processing behavior. Some students also engaged in using the cursor to highlight sentences as they read, which appeared to be a tactic to keep track of position in the text, as the highlighting was not permanent.

Text-processing behaviors unique to the audio condition included gaze patterns, playback speed adjustments, forwarding or rewinding audio clips, and fidgeting. Students gazed off the screen and particularly looked sideways (55.42) while listening, more than looking downward (37.20) or at the screen (13.16). Some students increased the audio playback speed while listening, a behavior that is reportedly still fringe among podcast and audiobook users despite its availability on audio player apps (Morris & Patterson, 2015; Roland et al., 2021).

A key behavior that was observed while students listened was fidgeting. *Fidgeting*, as defined in this study, are small physical movements or gestures that are often unconscious and sustained over time, for example playing with a pen or object. Fidgeting might be a physical manifestation of an effort to maintain *vigilant attention*, a term to delineate sustained attention to non-arousing tasks, by providing compensatory sensory input (Andrade, 2010; Langner & Eickhoff, 2013; Robertson & Garavan, 2004).

Fidgeting could also constitute a manifestation of mind wandering or inattentiveness (Carriere et al., 2013; Farley et al., 2013). This behavior, unlike kinesics, was only observed in the audio condition.

Table 5

Means, Standard Deviations, and Paired Sample T-tests of Text Processing Behaviors in Print and Audio

Variable	Print		Audio			
	M	*SD*	*M*	*SD*	*t*	*p*
Reading aloud duration (sec)	488.04	89.80	—	—	—	—
Duration (sec)	334.41	160.75	484.83	56.62	-10.23	<.001
Backward scrolling	1	1.64	—	—		—
Cursor reading (line-by-line)	0.99	3.30	—	—		—
Cursor reading (word-by-word)	1.75	10.36	—	—		—
Forward scrolling	25.67	16.45	—	—		—
Highlighting text	0.34	1.30	—	—		—
Screen stationary	55.52	46.18	—	—		—
Off-task behaviors						
Eating/drinking	0.45	1.58	0.86	1.39	-2.47	.007
Kinesics	0.70	1.02	2.27	2.03	-8.87	<.001
Yawning	0.23	0.58	0.56	1.08	-3.61	<.001
Gaze patterns						
Downward	—	—	37.20	36.56	—	—
Sideway			55.43	36.31		
On-screen			13.16	21.50		
Increase playback speed	—	—	0.15	0.56	—	—
Fidgeting	—	—	11.38	25.61	—	—
Forward audio	—	—	0.01	0.09	—	—
Rewind audio	—	—	0.17	0.64	—	—

Other audio text-processing behaviors included skipping ahead or re-listening to a section of the audio clip, both of which were not very common, as shown by the associated low mean values.

Post Reading and Listening

As shown in Table 6, print had an advantage over audio for constructed response items whether they assessed participants' ability to describe the main idea (gist understanding) or provide definitions for specialized vocabulary items for the same words that made up the topic knowledge pretest items. In contrast, for selected response items, there was no statistically significant difference between print and audio conditions regardless of whether the items assessed recall or inference. This finding is at odds with studies that have found a clear advantage in favor of print over audio on comprehension, even when assessed with selected response items (Singh & Alexander, 2022).

Table 6

Means, Standard Deviations, and Paired Sample T-tests of Scores on Comprehension Measures and Judgment of Performance Measures for Print and Audio

Variable	Print		Audio		*t*	*p*
	M	*SD*	*M*	*SD*		
Recall	0.69	0.19	0.68	0.20	0.39	0.35
Inference	0.64	0.21	0.62	0.21	0.64	0.26
Main idea	0.55	0.43	0.46	0.37	1.94	0.02
Vocabulary gain	0.13	0.30	0.05	0.26	2.19	0.01
Judgment of performance	0.56	0.20	0.47	0.21	5.89	<.001
Calibration of performance	-0.04	0.24	-0.12	0.25	3.20	<.001

Findings related to performance on comprehension measures assessing main ideas and recall of details are mixed in the medium research where text processing is compared across reading on paper and digitally. Some researchers have found that reading on paper

leads to better outcomes for remembering details and equal performance on main idea items pointing to the idea that remembering details requires more effortful processing than recalling the main idea (Singer & Alexander, 2018). At the same time, others have noted that reading on paper is related to better performance on main idea items and gives no advantage for recalling key points (Ronconi et al., 2022). In this study, I also found that reading print digitally or listening to text led to similar recall and inference comprehension outcomes and varied on the main idea or gist understanding performance. Several plausible explanations can be offered for this finding.

First, the text used in this study, *Enlightenment Now,* forwards arguments supported by scholarly literature and statistics. The reader or listener needs to follow those lines of support carefully and integrate them to arrive at the correct main idea. This could explain why inferring the main idea was more cognitively demanding in this study than in prior investigations. Second, the comparison in this investigation was between reading on screen and listening to audio. The comprehension performance results may have been different if listening to audio was compared to reading on paper. Finally, it is worth noting that students performed equally on recall and inference, better for main idea and vocabulary gain while reading than listening despite spending significantly lesser time reading (see Table 5 and Table 6).

The results related to judgment of performance and calibration (difference between perceived and actual performance) showed that students tended to be underconfident in both mediums (Table 6). On average, students were almost perfectly calibrated for print (-0.04) and underestimated their performance in the audio condition to a greater extent (-0.12). This finding, too, contrasts with what has been established in the

text processing and medium literature. Researchers have found that students overestimate their performance overall and even more when reading digitally than reading on paper (Ackerman & Goldsmith, 2011; Singer & Alexander, 2016, Vidal-Abarca et al., 2010). I expected a similar trend in the calibration results in the current study. However, students were not only close to 0 in calibration (i.e., perfectly calibrated), they tended to be underconfident. Previous studies have found accurate calibration to be associated with better self-regulation and comprehension (Bol & Hacker, 2012; Dunlosky & Rawson, 2012). But in this investigation, the underconfidence could be simply due to the higher complexity of the text than the ones used in previous studies.

Test-processing behaviors were captured through Qualtrics® (test-taking duration) and Zoom® video (revisions and reframing). As shown in Table 7, there was no difference in the time taken to complete the different components of the comprehension measure between print and audio. However, on average, participants revised their responses to the recall part of the comprehension measure a significantly greater number of times after reading in print than listening to audio (0.98 vs. 0.57). This revision pattern could indicate higher metacognitive monitoring facilitated by print than audio (Couchman et al., 2016). No other significant difference was detected in revision or reframing frequencies.

Post Task

As displayed in Table 8, participants reported enjoying reading print significantly higher than listening to audio. They also perceived audio to be significantly more difficult, a perception borne out by their inferior comprehension performance in audio, at

Table 7

Means, Standard Deviations, Paired Sample T-tests of Test Processing Behaviors in Print and Audio

Variable	Print		Audio		*t*	*p*
	M	*SD*	*M*	*SD*		
Selected response duration						
Recall	59.96	34.58	60.07	31.78	-0.04	0.49
Inference	59.83	34.93	58.74	30.19	0.54	0.29
Constructed response duration	139.22	100.18	133.08	108.92	0.67	0.25
Revise selected response						
Recall	0.98	1.48	0.57	0.98	2.77	0.003
Inference	0.85	1.33	0.73	1.30	0.83	0.20
Reframe						
Main idea	0.27	0.57	0.35	0.63	-1.12	0.13
Vocabulary	0.20	0.49	0.15	0.47	0.85	0.19

least in constructed response items (see Table 6). The scale for these items ranged from 0 to 100, where 0 was not enjoyable or difficult at all, and 100 was very enjoyable or difficult. Overall, enjoyment levels for the task stood at a middling 53%, and difficulty levels were, on average, at 60% for the task.

Table 8

Means, Standard Deviations, Paired Sample T-tests of Self-Reported Task Experience for Print and Audio

	Print		Audio		*t*	*p*
	M	*SD*	*M*	*SD*		
Enjoyment	0.59	0.23	0.47	0.26	6.97	<.001
Difficulty	0.58	0.22	0.62	0.23	-2.29	.01

Research Question 2a: Reader and Listener Profiles

The second research question aimed to explore how learner characteristics, text, and test processing behaviors characterize reader and listener profiles. The print and

audio mixture models included only informative variables related to the learner, task, and test. For example, self-efficacy was dropped as an indicator variable for reader profiles because it had limited variability and did not differentiate the profiles. Likewise, very few students forwarded the audio clip; therefore, the variable was not included when constructing audio profiles. I compared models by extracting two to five profiles for print and audio.

The models were compared across four covariance structures (Bauer & Curran, 2004; Pastor et al., 2007). In Model A, which represents the latent profile analysis (LPA) model, indicator variances were constrained to be equal across profiles, and covariances were fixed at zero. In Model B, indicator variances were estimated freely across profiles while the covariances remained fixed at zero. In Model C, indicator variances were freely estimated across profiles, and covariances were constrained to equality across profiles. Finally, in Model D, both indicator variances and covariances were freely estimated across profiles (see Table 9a and Table 9b).

Several sources of information were considered to select the final models as no single statistic has been unequivocally shown to be preferable (Nylund et al., 2007; Tien et al., 2013). Specifically, I considered model fit indices, such as the Bayesian information criterion (BIC) and the bootstrap likelihood ratio test (BLRT). Lower BIC values indicate better fit. The BLRT compares models with n and n-1 profiles. Significant BLRT value is evidence in favor of the model with n profiles compared with n-1 profiles (Pastor & Erbacher, 2018). I also considered conceptual information and the size of the profiles in selecting the best-fitting models for print and audio. Due to computational

Table 9a

Reader Profiles: Print Mixture Models A-D

Profiles	Model	Covariance structure	AIC	BIC	aBIC	BLRT	EN
2 (1,126)	A	Fixed variance, zero covariance	11267.31	11346.95	11258.4	-5687.94* (0.0000)	1
2 (87,40)	B	Free variance, zero covariance	10913.22	11018.46	10901.447	-5687.94* (0.0000)	0.938
2 (78,49)	C	Free variance, fixed covariance	10692.56	10900.19	10669.328	5538.17* (0.0000)	0.95
2 (80,47)	D	Free variance, free covariance	10705.54	11015.55	10670.848	-5538.17* (0.0000)	0.964
3 (2,110,15)	A	Fixed variance, zero covariance	11159.78	11267.85	11147.682	-5610.31* (0.0000)	0.958
3 (66,17,44)	B	Free variance, zero covariance	10765.11	10924.39	10747.29	-5419.61* (0.0000)	0.954
3 (68,45,14)	**C**	**Free variance, fixed covariance**	**10527.36**	**10789.02**	**10498.077**	**-5289.16* (0.0000)**	**0.927**
3 (65,20,42)	D	Free variance, free covariance	10571.57	11038.02	10519.381	Could not be computed	0.949
4 (101,9,15,2)	A	Fixed variance, zero covariance	11061.4	11197.92	11046.12	-5537.68* (0.0000)	0.97
4 (59,16,8,42)	B	Free variance, zero covariance	10643.41	10856.72	10619.542	-5403.07* (0.0000)	0.943
4 (62,11,43,11)	C	Free variance, fixed covariance	10479.32	10795.02	10443.993	Could not be computed	0.952
4 (25,61,23,18)	D	Free variance, free covariance	10623.86	11246.74	10554.168	-5164.5* (0.0000)	0.963
5 (3,6,2,101,15)	A	Fixed variance, zero covariance	10979.61	11144.57	10961.148	-5475.85* (0.0000)	0.976
5 (48,17,43,2, 17)	B	Free variance, zero covariance	10611	10878.35	10581.085	-5265.28* (0.0000)	0.928
5 (71,10,16,10, 20)	C	Free variance, fixed covariance	10427.49	10797.23	10386.115	Could not be computed	0.938
5	D	Free variance, free covariance	Model could not be estimated.				

Note. p-value given in parenthesis. Selected model in bold.

Table 9b

Listener Profiles: Audio Mixture Models A-D

Profiles	Model	Covariance structure	AIC	BIC	aBIC	BLRT	EN
2 (122,5)	A	Fixed variance, zero covariance	14666.934	14871.716	14644.021	-7356.762* (0.0000)	1
2 (41,86)	B	Free variance, zero covariance	14184.278	14420.345	14157.864	-7356.762* (0.0000)	.95
2 (89,38)	C	Free variance, fixed covariance	14087.349	14479.847	14043.432	Could not be computed	.971
2 (51,76)	D	Free variance, free covariance	14037.837	14586.765	13976.417	-7244.568* (0.0000)	.988
3 (44,53,30)	A	Fixed variance, zero covariance	14696.172	14989.124	14663.393	-7288.152* (0.0000)	.778
3 (38,75,14)	**B**	**Free variance, zero covariance**	**14177.299**	**14532.823**	**14137.519**	**-6976.080* (0.0000)**	**.925**
3 (43,77,7)	C	Free variance, fixed covariance	14031.340	14543.293	13974.056	-6916.992* (0.0000)	.976
3 (127,0,0)	D	Free variance, free covariance	15069.135	15893.949	14976.845	-6961.344 (1.0000)	1
4 (3,23,100,1)	A	Fixed variance, zero covariance	14635.806	15016.927	14593.162	-7217.428* (0.0000)	.994
4 (93,34,0 0)	B	Free variance, zero covariance	14549.635	15024.614	14496.488	-7107.817 (1.0000)	.954
4	C	Free variance, fixed covariance	Model could not be estimated.				
4 (55,72,0,0)	D	Free variance, free covariance	14646.092	15746.792	14522.932	-6956.366 (0.2413)	.993
5 (15,25,79,5, 3)	A	Fixed variance, zero covariance	14673.595	15142.886	14621.085	-7151.210 (1.0000)	.903
5 (59,0,68,0,0)	B	Free variance, zero covariance	14648.888	15243.324	14582.376	-7096.144 (1.0000)	.967
5	C	Free variance, fixed covariance	Model could not be estimated.				
5 (0,0,127,0,0)	D	Free variance, free covariance	15457.135	16833.722	15303.106	-7244.568 (1.0000)	1

Note. p-value given in parenthesis. Selected model in bold.

limitations, some models did not converge. I could not increase the random starts to a high enough value and run the models successfully.

Reader Profiles

The final model selected for readers consisted of three profiles based on a combination of statistical and theoretical considerations (Model 3C; see Table 9a). Statistically, I examined BIC and BLRT to aid in the selection of the best-fitting model. Lower BIC values indicate better fit. The BLRT compares models with n and n-1 profiles. Significant BLRT value is evidence in favor of the model with n profiles compared with n-1 profiles (Pastor & Erbacher, 2018).

Theoretically, I considered the substantive interpretability of the profiles, as it is important to ensure that the profiles are meaningful and useful in understanding the readers. The three profiles were found to be statistically and substantively interpretable, with clear differences between the profiles in terms of reading habits, text processing, and test processing behaviors.

Additionally, I examined a model with four reader profiles (Model 4c) because its goodness-of-fit statistics appeared to be promising. This model had similar profiles to the three-profile model, with a few readers from each profile forming a fourth group. However, upon further analysis, I found that this fourth profile did not add any substantive value to understanding the characteristics of the readers. Therefore, I opted to use the model with three profiles for our subsequent analysis and interpretation.

The profile assignment probabilities for the three groups were 0.97, 0.97, and 1, respectively. For readers, the profiles differed in both shape and level (Figure 10). I examined the means for each indicator variable across the profiles to interpret the reader profiles (see Table 10). My examination of the pattern of means revealed the following profiles: (a) Distracted Surfers (n = 68), (b) Labored Harvesters (n = 45). and (c) Fluent

Surveyors (n = 14). I named the profiles based on their most defining characteristic to aid the following discussion.

Reader Profile 1: Distracted Surfers. This profile constituted the largest student grouping (n = 68). Concerning learner characteristics, the students in this profile possessed the least number of books and eBooks. They reported reading approximately 13 hours per week. Their text-processing behaviors were characterized by their propensity to get through the experimental text fastest (279 s) with a comparatively higher forward scroll frequency (23.8).

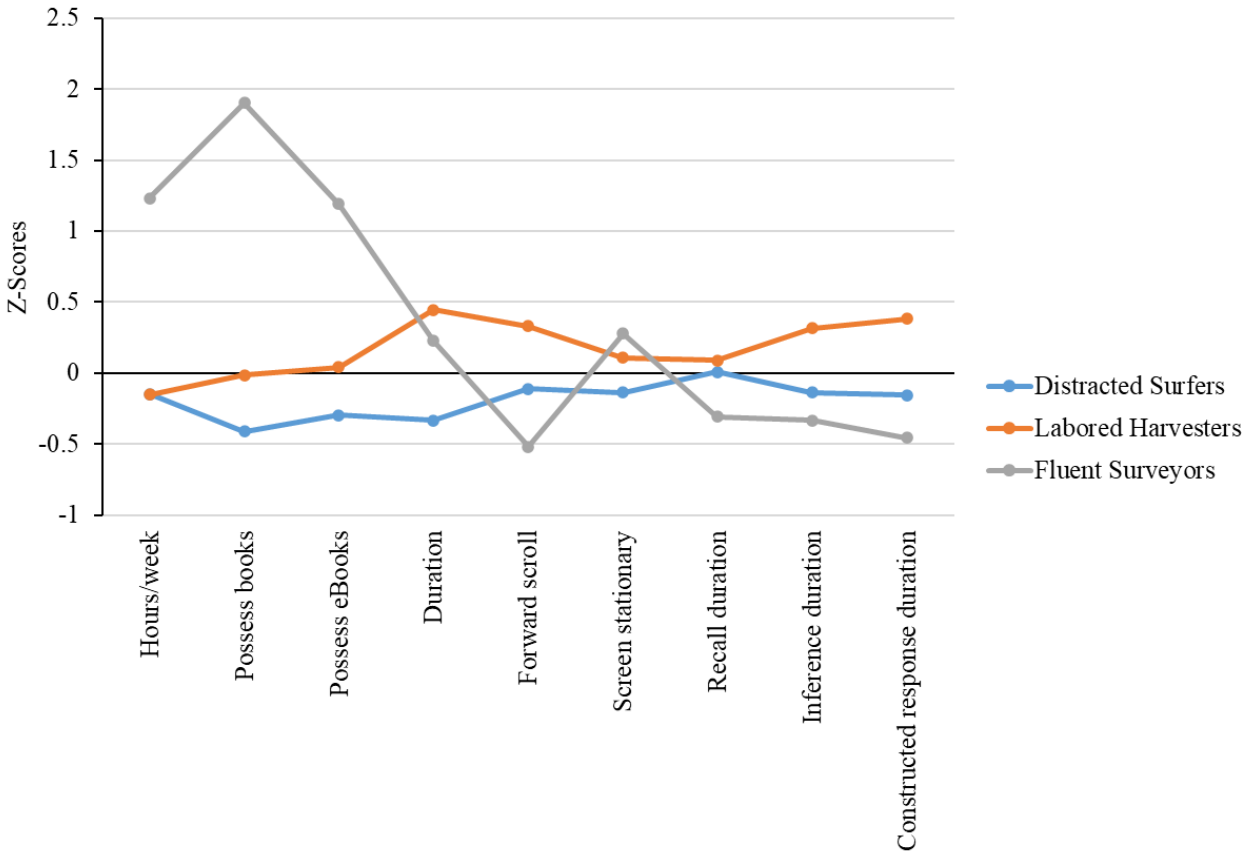
The Distracted Surfers also spent the least time dwelling on the textual content. This lack of reflection was captured in the indicator variable labeled "stationary." This variable measured the periods without any scrolling (see Table 10). Given the complex nature of the text, it was critical to process each sentence to draw an integrated mental model of the text's main message. Probably as a consequence of not spending adequate time on the text to construct meaning, the Distracted Surfers took longer to complete the tests. They were in the middle on test duration across all comprehension measures among the three reader profiles. They likely struggled to arrive at the correct responses because they were not unsure of the surface-level details, inference-based understanding, and main idea communicated in the text.

Reader Profile 2: Labored Harvesters. The Labored Harvesters comprised the second-largest reader group (n = 45). This profile was characterized by taking the longest to read among the three groups (404.2 s). The Labored Harvesters also scrolled forward the most (31.14). They also took the longest time among the three profiles to complete all the comprehension measures. Hence, the descriptor "Labored." They reported possessing

more books and eBooks (approximately 30) than the Distracted Surfers but read about the same amount weekly. Their text-processing and test-processing behaviors suggested more of an information-managing approach and test-taking orientation. It appears that the Labored Harvesters were trying to anticipate what could be on the test and looking for (or harvesting) information that they deemed the most important (Alexander, 2018).

Figure 11

Comparison of Reader Profiles across Indicator Variables



Reader Profile 3: Fluent Surveyors. The Fluent Surveyors made up the smallest profile group (n = 14). They self-reported having the largest collections of books and devoting the most time to reading compared with the other two profile groups. The Fluent

Surveyors spent, on average, 25.76 hours per week reading and had approximately 96 books and 25 eBooks in their collection.

Regarding text-processing behaviors, the Fluent Surveyors took longer to read the experimental text (368.81 s) than the Distracted Surfers but lesser than the Labored Harvesters. They made the fewest forward scrolls (17) and held the screen stationary the longest among the three profile groups. This suggests that they took a deliberate and reflective approach to text comprehension. The strategy aligns with the complexity of the text, which requires intratextual integration for successful comprehension.

The deep-processing approach taken by these fluent readers was also evidenced in the short time they spent completing the comprehension questions. It appears they were sure of their responses and did not require as much time as the other profile groups to respond. The 78.s s average duration for constructed response items, compared with 118.5 s and 189.9 s for the Distracted Surfers and Labored Harvesters, respectively, is also indicative of a facility with writing. This facility with writing could also be linked to their strong reading habits (Fitzgerald & Shanahan, 2000; Shanahan, 2016). See Table 10 for mean scores on indicator variables.

Table 10

Means of Indicator Variables for Reader Profiles

Indicator Variables	Distracted Surfers (*n*=68)		Labored Harvesters (*n*=45)		Fluent Surveyors (*n*=14)	
	Mean	SE	Mean	SE	Mean	SE
Learner						
Hours read per week	13.128	1.046	13.099	0.875	25.757	3.881
Possess Books	7.271	0.92	22.454	2.453	95.768	18.665
Possess eBooks	2.119	0.268	7.253	0.92	24.66	10.288
Task						
Duration (sec)	279	14.908	404.207	28.566	368.811	45.682
Forward scrolling	23.825	2.16	31.135	3.131	17.15	2.978

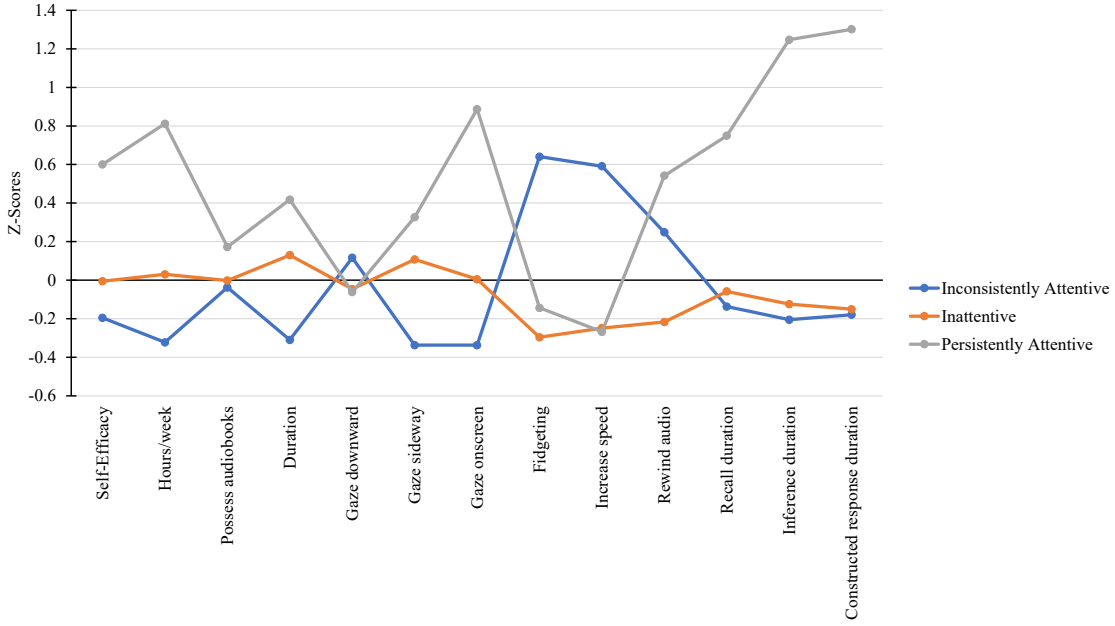
Screen stationary	49.251	4.681	60.502	8.132	68.46	15.089
Test						
Selected response duration (sec)						
Recall	60.265	3.549	63.072	5.957	49.395	8.806
Inference	55.036	3.549	70.888	7.13	48.131	7.37
Constructed response duration (sec)	118.481	8.924	189.869	22.358	78.816	9.074

Listener Profiles

The final model for listeners consisted of three profiles (Model 3B; see Table 9b). The profile assignment probabilities for the three groups were 0.96, 0.98, and 0.95. Similar to the reader profiles, the listener profiles also differed in shape and level (Figure 11). I examined the pattern of means across the three profile groups to ascertain their most salient characteristics and assign them names (Table 11). Subsequently, I derived the following listener profiles: (a) Inconsistently Attentive (n = 38), (b) Inattentive (n = 75), and (c) Persistently Attentive (n = 14).

Figure 12

Comparison of Listener Profiles across Indicator Variables



Listener Profile 1: Inconsistently Attentive. The Inconsistently Attentive profile was the second largest grouping of students (n = 38). This group included those who fidgeted more than the ones in other profiles (27.8 times on average compared with 3.8 and 7.7). This behavior may be indicative of a strategy for managing attention (Langner & Eickhoff, 2013; Robertson & Garavan, 2004). They also possessed the fewest audiobooks, with an average of less than 1, pointing to their low experience with the audio medium. Despite being the least experienced with audio, they made the playback speed faster more often than those in other groups, a behavior typically associated with those who listen to podcasts more often (Morris & Patterson, 2015).

Consequently, those populating the Inconsistently Attentive profile were the fastest at listening to the experimental audiobook excerpt at 465.55 s. On average, students in this group gazed down more than those in other profiles. Observationally, it appeared that often when students looked downward they were on their phones. Unfortunately, because the phones were not visible on the recorded screen, I could not be

Table 11

Means of Indicator Variables for Listener Profiles

Indicator Variables	Inconsistently Attentive (n=38)		Inattentive (n=75)		Persistently Attentive (n=14)	
	Mean	SE	Mean	SE	Mean	SE
Learner						
Self-Efficacy	0.711	0.021	0.741	0.021	0.837	0.035
Hours listened per week	3.121	0.445	5.506	0.847	10.788	4.044
Possess audiobooks	0.907	0.370	1.014	0.417	1.534	0.879
Task						
Duration (sec)	465.546	15.342	490.442	0.648	506.740	9.218
Gaze patterns						
Downward	41.462	7.496	35.520	4.363	34.933	7.948
Sideway	43.194	5.107	59.341	5.014	67.289	14.514
On screen	5.913	1.440	13.270	3.738	32.238	9.127
Fidgeting	27.789	6.964	3.793	1.044	7.711	3.430
Increase playback speed	0.481	.035	0.011	0.000	0.000	0.000
Rewind audio	0.329	0.04	0.031	0.000	0.517	0.02
Test						

Selected response duration (sec)						
Recall	55.710	5.577	58.212	4.057	83.889	13.241
Inference	52.549	5.484	54.983	4.834	96.408	11.811
Constructed response duration (sec)	113.554	14.857	116.661	12.644	274.833	66.597

sure of the behavior. They were also quickest at completing the comprehension measures. The students in this profile appeared to be the most interested in getting the task done as quickly as possible.

Listener Profile 2: Inattentive. The majority of students populated this profile group (n = 75). They were in the middle of the other two profile groups with respect to hours devoted to podcasts and audiobooks weekly (5.5 hr) and possession of audiobooks (1.01).

Concerning text processing behaviors, those in the Inattentive profile again stood as the group in the middle. They took more time to listen (490.4 s) than those in the Inconsistently Attentive profile but lesser than those in the Persistently Attentive group. This is because they were not inclined to increase the audio speed and the least into rewinding audio when compared to those in the other two profiles (see Table 11 and Figure 11).

In keeping with this middling trend, they gazed downward (35.52), less than the Inconsistently Attentive profile students but more than the Persistently Attentive individuals. Likewise, they looked sideways (59.34) more often than the first profile but less than the third one. The same pattern was observed for looking directly at the screen (13.27). This profile group appears to constitute students who were passive listeners when we put together their fidgeting behavior (or lack thereof), their gazing patterns, and the time spent listening to the experimental audiobook text,

Students in the Inattentive profile were also in the middle of the two profiles on test-processing behaviors. They took slightly more time to complete the selected and constructed-response items than the Inconsistently Attentive profile students but lesser than the Persistently Attentive ones.

Listener Profile 3: Persistently Attentive. This profile was characterized by its unique gazing patterns. Of the three profiles, they looked downward (34.9) the least but gazed sideways (67.3) the most. The students who constituted this profile also looked directly at the screen, perhaps to track the progress of the audiobook clip.

The students in the Persistently Attentive profile had greater experience with audiobooks and podcasts than other profile groups. Those who made up this profile reported spending over 10 hours per week listening and possessed the most audiobooks. They took the longest to listen to the experimental audiobook clip (506.7s), perhaps because they also re-listened and did not increase the playback speed. It appears that since they had more experience with audio, they were able to monitor their comprehension and took their time due to the complexity of the text.

In terms of test processing, this profile was also the slowest in completing all the different sections of the comprehension measure (83.9 s for recall, 96.4 for inference, and 274.83 for main idea and vocabulary). They were nearly twice as slow when compared with other profiles. Without considering the performance on the comprehension measures, it is difficult to ascertain whether the highest test-taking duration was attributable to more uncertainty or better metacognitive monitoring (Couchman et al., 2016).

Research Questions 2b and 2c: Validation of Reader and Listener Profiles

The goal of RQ2b and RQ2c was to establish the construct validity of the identified profiles by investigating whether the profiles show meaningful and well-differentiated relations to key external (auxiliary) variables. Comprehension scores on recall, inference, vocabulary gain, main idea (gist understanding) items, and calibration of performance scores were entered as the distal outcome variables using the BCH method in *MPlus© 8.8* (Bolck et al., 2004; Muthén & Muthén, 1998-2017). This method accounts for the classification error for each individual (Nylund et al., 2013).

Topic knowledge was set as the covariate and used to predict profile membership based on theoretical relations between prior knowledge and text processing (Afflerbach, 1986; Kintsch, 1988). This analysis was carried out with the automated R3STEP procedure in *MPlus© 8.8,* which uses logistic regression to predict profile membership using covariates.

Research Question 2b: Comprehension Outcomes Across Profiles

Reader Profiles. Statistically significant differences in comprehension were observed for the recall and main idea items (see Table 12). In particular, pairwise contrasts showed that students in the Labored Harvester profile scored the highest on recall items (0.73). Their recall comprehension scores were significantly higher than those in the Distracted Surfers profile ($\chi^2(1) = 4.443$, $p = 0.03$) but similar to those in the Fluent Surveyors group.

With regard to main idea performance, individuals in the Fluent Surveyor profile scored highest (0.68) despite spending the least time on the constructed response items among the three profile groups (see Figure 10). They differed significantly only from the Distracted Surfers ($\chi^2(1) = 4.65$, $p = 0.014$). Still, no statistical difference was detected

between the students in the Labored Harvester and Fluent Surveyor profiles on the main idea item.

Based on the comprehension performance scores it appears that those in the Labored Harvester group benefit in terms of comprehension from their effortful engagement with the text and the test items. They seem to compensate successfully for their lower reading experience by investing more time on task.

The three profile groups showed no statistical differences on the remaining comprehension measures or on calibration accuracy.

Table 12

Comparison of Reader Profiles on Distal Outcomes

	Distracted Surfers (*n*=68)		Labored Harvesters (*n*=45)		Fluent Surveyors (*n*=14)		Overall χ^2
	M	*SE*	*M*	*SE*	*M*	*SE*	
Recall	0.661_{a}	0.024	0.735_{b}	0.025	0.685_{ab}	0.063	4.457
Inference	0.647	0.025	0.651	0.035	0.563	0.061	0.418
Main idea	0.504_{c}	0.043	0.612_{cd}	0.064	0.679_{d}	0.018	4.495
Vocabulary gain	0.125	0.035	0.115	0.052	0.214	0.079	1.212
Calibration	-0.034	0.029	-0.070	0.038	0.026	0.058	1.918

Note. Means with different subscripts differ at the $p = .05$ level for each row displaying scores on different comprehension and calibration outcomes. *p<.05

Listener Profiles. Overall, the listener profiles differed significantly on main idea ($\chi^2(2) = 13.334$, $p = 0.001$) and calibration of performance ($\chi^2(1) = 5.806$, $p = 0.05$). Pairwise contrasts showed that those in the Inconsistently Attentive and Persistently Attentive profile outperformed those in the Inattentive profile on main idea ($\chi^2(1) = 5.239$, $p = 0.02$; $\chi^2(2) = 11.233$, $p = 0.001$) but students in the Inattentive group were

better calibrated than the ones in the Inconsistently Attentive profile ($\chi^2(1) = 5.806$, $p = 0.016$). Although the omnibus chi-square test was non-significant for vocabulary gain, the Persistently Attentives outperformed the Inconsistently Attentives ($\chi^2(1) = 4.603$, $p = 0.032$). Table 13 presents a comparison of the scores on the distal outcomes of interest and results from the equality of means test conducted using the BCH procedure. As shown by the mean scores, the three profiles performed similarly on recall and inference components of comprehension, which were also the selected-response items. The difference in scores on only the constructed response items, may be an indication that individual-level differences and processing behaviors only influence performance where higher-order processes are involved, such as integration.

Table 13

Comparison of Listener Profiles on Distal Outcomes

	Inconsistently Attentive (*n*=38)		Inattentive (*n*=75)		Persistently Attentive (*n*=14)		Overall χ^2
	M	*SD*	*M*	*SD*	*M*	*SD*	
Recall	0.72	0.006	0.67	0.003	0.67	0.010	1.792
Inference	0.64	0.006	0.61	0.003	0.67	0.015	1.410
Main idea	0.54_a	0.010	0.37_b	0.005	0.70_a	0.023	13.334*
Vocabulary gain	0.01_c	0.006	0.04_{cd}	0.004	0.19_d	0.020	4.605
Calibration	-0.19_e	0.006	-0.07_f	0.004	-0.13_{ef}	0.018	5.806*

Note. Means with different subscripts differ at the $p = .05$ level for each row displaying scores on different comprehension and calibration outcomes. *$p<.05$

Research Question 2c: Prior Knowledge to Predict Profile Membership

To establish the validity of the profiles, I also included prior knowledge as a covariate to check if prior topic knowledge predicted profile membership. Similar to the analysis conducted with the distal outcomes, wherein the variables were included in the

model itself, the covariate too was entered as an auxiliary variable once the final models had been selected using the 3-step procedure in *MPlus©.*

Reader Profiles. The results from the multinomial logistic regression showed that prior topic knowledge significantly predicted membership to the Labored Harvester profile (see Table 14). Specifically, higher levels of prior topic knowledge were associated with a greater likelihood of belonging to the Labored Harvester profile in comparison with the lowest performing profile, the Distracted Surfers (p = 0.026).

To make the interpretation of the coefficients more intuitive, the odds ratios for the prior topic knowledge covariate are also presented in Table 14. For each unit increase in prior topic knowledge, the odds of belonging to Labored Harvester profile increased by a factor of 9.69, which was statistically significant. Further, with each unit increase in prior knowledge, the odds of belonging to the Fluent Surveyor profile increased by a factor 22.04 in comparison with the odds of belonging to the Distracted Surfer Profile. However, the odds ratio associated with the Fluent Surveyor profile was non-significant.

The high odds ratio suggests that the prior knowledge has a potentially important effect on the membership to the Fluent Surveyor profile, even if the effect was not statistically significant. The magnitude of the odds ratio can be interpreted as the effect size, which represents the strength of the relationship between the predictor and outcome variables. It is important to note that a non-significant result does not necessarily mean that there is no effect. Rather, it may simply reflect a lack of statistical power to detect the effect with the sample size used in this study. Future research with a larger sample size may be able to detect a significant relationship between prior knowledge and membership to the Fluent Surveyor profile.

Table 14

Logistic Regression Results with Distracted Surfers as Reference Profile

Effect	Estimate	SE	*p*	Odds ratio
Intercept				
Labored Harvester	-1.697	0.640	0.008	—
Fluent Surveyor	-3.366	1.104	0.002	
Beta coefficient				
Labored Harvester	2.272	1.020	0.026	9.696*
Fluent Surveyor	3.092	1.642	0.060	22.014

*$p<.05$

Listener Profiles. Results from the multinomial logistic regression indicated that prior topic knowledge is a significant predictor to the Persistently Attentive profile (see Table 15). In particular, higher levels of prior topic knowledge were associated with a greater likelihood of belonging to the Persistently Attentive profile in comparison with the lowest performing group, the Inattentive profile ($p = 0.048$). For each unit increase in prior topic knowledge, the odds of belonging to the Inconsistently Attentive profile were 3.26, and the odds of membership to the Persistently Attentive group were 9.7 in comparison to membership to the Inattentive profile (see Table 15).

Based on these results, it appears that topic knowledge is a predictor of profile membership only for the group with the highest comprehension performance.

Table 15

Logistic Regression Results with Inattentive as Reference Profile

Effect	Estimate	SE	*p*	Odds ratio
Intercept				
Inconsistently Attentive	-1.377	0.641	0.032	—
Persistently Attentive	-3.058	0.781	0.000	
Beta coefficient				
Inconsistently Attentive	1.183	1.024	0.248	3.265
Persistently Attentive	2.272	1.148	0.048	9.701*

*$p<.05$

Research Questions 3a and 3b: Mixed Methods Integrative Results

The final set of research questions aimed to integrate the results from the quantitative and qualitative analysis. Selected members from each profile group were interviewed. The data were coded based on the four dimensions represented by the tetrahedral model (learner, text, task, and test), which also guided the quantitative profiling. Following the same order as the first two sets of research questions, the qualitative results are first presented, indexed by the two mediums. Next, the qualitative findings are given for each reader and listener profile to explicate further the nature of those populating the profiles, which were unearthed by the quantitative finite mixture modeling analysis. Thusly, the qualitative results are discussed in conjunction with the primary quantitative findings.

Research Question 3a: Qualitative Differences Across Mediums

The frequency counts for the learner, text, task, and test dimensions that emerged in participants' interviews for the print and audio mediums are displayed in Table 16. In the interviews, participants discussed more for audio compared to the print condition, particularly for the learner dimension. When it came to the text dimension, the interviewees shared more about the print medium than the audio medium. This higher number of text-related utterances for the print portion might explain the superiority of print for understanding the main idea since print seems to facilitate an awareness of the vocabulary, tone, and themes presented in the text.

The number of references to the task or test dimensions across mediums was very similar.

Learner-Related Utterances

The utterances related to the learner dimension were further divided into *affect, habits and preferences, metacognition, self-efficacy*, and *topic or prior knowledge*. The frequency of affect codes for print and audio mediums suggests that listening invoked more affective verbalizations than reading (see Table 16). Besides the difference in counts, there was also a divergence in the valence of the expressed ideas. Specifically, in print, there were some positive or mildly negative emotions expressed (e.g., "I think I was feeling more confident" and "slightly frustrated"). At the same time, in audio the affect were mostly negative with emotions ranging from boredom to annoyance (e.g. "In the beginning I just was kind of bored" and "I find it audio to be bit more annoying"). This corresponds to how the students reported finding print more enjoyable and less difficult than audio in the post-task measure (see Table 8).

Likewise, for the habits and preferences code, audio had higher frequency than print. The messages ranged from preferring reading on screen versus paper (e.g., "reading on screen specifically does not bother me as like versus reading in print) to the opposite of disliking screen (e.g., "I do not like to read any content on the screen"). It seems that listening to audio in the context of this study made students recollect online instruction during COVID-19 lockdown. On the other hand, all but one interviewee expressed something negative, from not being a "podcast person" to not enjoying "watching videos for class." The only student who professed a preference for audio talked about it as a "fantastic way for me to know certain information." It is worth noting that this participant was an English foreign language learner. The qualitative results again converge with the quantitative results, where participants reported spending more time reading weekly than

listening. The qualitative interviews tapped into a preference for reading on screen rather than reading on paper for some students. This preference was not captured in the quantitative measure, which did not include an item for relative medium preferences.

The metacognition code was also attributed less frequently to utterances in print than audio. Interestingly, many of the idea units under metacognition in print were about what participants would have liked to do but did not. One of the participants expressed how they would have liked to use the methods learned in their "English class" to check "credibility." Another interviewee wanted to "take notes." In the audio, there was only one person who regretted not "having a pad to write down notes," while the others talked about being "distracted" and not paying attention.

Self-efficacy had the lowest frequency counts in both print and audio mediums. The higher self-efficacy for reading over listening corroborated with the higher print self-efficacy reported in the quantitative measure (see Table 4 for quantitative results pertaining to pre-reading and listening measures). In print, an interviewee extolled their ability to read fast, and in audio, someone talked about being able to read well and, therefore, not choosing audio. They mentioned that audio was seen as support for struggling readers in their school. The ideas expressed coded for topic or prior knowledge in print and audio were similar. Participants talked about being "familiar" with the principles or coming across them in their "philosophy class."

Text-Related Utterances

Similar to the learner dimension, the utterances associated with the text dimension were further coded for: *affect, structure, themes/ideas, tone, syntax*, and *vocabulary* (see Table 16). With regard to affect, print and audio mediums had divergent reactions.

Whereas in print, the interviewees found the content interesting (e.g., “the article was very like interesting about that”), in audio, they expressed the opposite (e.g., “it was my least favorite section”)—corroborating the findings from the quantitative measure on perceived task enjoyment and difficulty across the two mediums. Concerning text structure, interviewees expressed similar issues for print and audio mediums. The most frequent complaint, irrespective of the medium, was that the text did not establish connections among ideas. This was encapsulated in a comment that compared the structure of the text to a “tangled web.”

A critical difference in the interviewees’ utterances related to print and audio had to do with prior knowledge activation. For example, when describing text encountered in the print medium, participants spoke of—“life expectancies,” “Marxism, communism,”—ideas that were not directly mentioned in the text. On the other hand, in audio, the discussion remained limited to broad themes explicitly stated in the text, such as meaning and happiness).

Further, the interviewees brought up specific details from the text seen in the print medium, even when the video segment did not include that particular segment (e.g., “I remember that one [Seems Sally story] was the one that bothered me the most”). The fact that students recalled precise details from what they read in the past suggests that there may be delayed memory effects related to mediums. Individuals may retain information longer after reading than listening to text.

With respect to tone, the salient aspect related to the print medium utterances was to see the text as “academic” and written for a “scholarly audience.” In audio, one of those interviewed felt it was “conversational,” while another thought it felt like a

"lecture." Unsurprisingly, none of the idea units expressed for the audio medium included references to the syntax. In the print medium, two of the participants verbalized that they found the sentences lengthy ("it has a lot of really long sentences"). A similar trend was observed with the vocabulary code. Only one idea unit in audio was tagged with the vocabulary code. In contrast, in print, nine idea units included references to vocabulary. In print, interviewees talked about not knowing the "meaning" of some words and also pointed out the use of "jargon."

Overall, the print medium evoked more verbalizations related to the text dimension than the audio medium. This imbalance between the two mediums may indicate that the affordances of the medium influence engagement with text. Reading print allows individuals to see the words, process sentences, and reflect on themes and ideas, while in audio, the ephemeral quality of the content precludes the possibility of engaging with the mechanics of the text (Rubery, 2016; Singer & Alexander, 2017; Surber & Schroeder, 2007).

Task-Related Utterances

Task-related utterances were coded for *elements/features, cognitive processing behaviors, physical processing behaviors*, and *task experience*. While the fewest task-related utterances included references to elements/features in the print medium, the reverse was true for the utterances related to the audio medium. In print, the three verbalizations concerning elements/features each captured a different notion. One compared the task to the Test of English as a Foreign Language (TOEFL®), and another lamented the lack of a "highlighting feature." At the same time, the final one celebrated the ability to "look at" and "interact with" something.

The most frequently expressed idea in the utterances regarding elements/features in audio was centered on the narrator. The second most popular verbalization was about the playback speed ("monotone," "robotic," "he was talking very slowly"). The other prominent notion was the lack of "physical" text in "front of you." This lack of physicality associated with the audio medium could be one explanation for why students engaged in more off-task behaviors in the audio medium versus the print medium (see Table 5 for text processing behaviors). More than twice the verbalizations in the print medium contained references to cognitive processing than the audio medium (see Table 16). In print, interviewees expressed a range of cognitive processing behaviors and strategies, such as looking for the main idea and supporting details, re-reading for clarifications, paraphrasing, summarizing, connecting to prior knowledge, and asking questions about the data presented in the text. One participant also mentioned using translation software to understand the text better.

In contrast, the range of cognitive behaviors was much more limited in the audio medium. Participants primarily tried to look for themes, made connections to prior knowledge, and attempted to determine the narrator's nationality. The preponderance of utterances related to cognitive processes in the print medium could explain why print had an advantage over audio in the main idea and vocabulary gain items.

Similar to cognitive processing, physical processing behaviors were also more diverse in print than in audio. Indeed, reading print by nature necessitates more task-related physical behaviors than listening to audio. In the print medium, participants detailed behaviors such as "scrolling back," "reading with the mouse," "skimming it and getting the information that I could," and "reading aloud." In contrast, in audio,

interviewees most frequently described off-task behaviors, such as “playing with my fingers,” “looking at my hands,” “play with my cat.”

The fourth code captured task experience. More utterances in the print medium included references to task experience than in the audio medium. The differences between the two mediums were even more pronounced in the quality of the idea units. The prevailing messages in print and audio were again influenced by the engagement afforded by the mediums as captured in the following comments: “doing something myself was a lot different experience than having something done read to me” and “feeling focused, engaged at least when I was doing the reading.” Several participants also mentioned that the reading task felt like it was for “school” or “class.” On the other hand, the audio task felt “novel” and, at the same time brought back (negative) memories of online lectures for some, encapsulated in this exclamation, “My God! I spent two years already doing this.”

Most students prefer reading in print over listening to audio, a quantitative result corroborated by the qualitative findings. Still, as expressed by some interviewees, the audio medium was preferable to non-native English speakers for processing texts in English.

Test-Related Utterances

The utterances related to the test dimension were segmented into *general* and *specific*. The frequency distribution for the general and specific codes was more balanced for the print than the audio medium (see Table 16). Interviewees verbalized mixed ideas in both the print and audio mediums without much consistency.

Within the utterances coded for *general,* some claimed they understood the content, while others talked about not comprehending. This was the case regardless of the mediums. The two participants who perceived that the audio medium was better for comprehension ("I understand more when I listen to this audiobook") did not have English as their first language, whereas those who complained about audio ("it was difficult for me") were English native speakers. These verbalizations may point to an opportunity for using audio for content delivery with those learning a second or foreign language.

Table 16

Learner, Text, Task, and Test Dimension Counts for Print and Audio Mediums

Dimension Subdimension	Print	Audio	Total
Learner	99	140	239
Affect	5	14	19
Habits and Preferences	8	26	34
Metacognition	80	94	174
Self-efficacy	1	2	4
Topic/Prior Knowledge	5	4	8
Text	50	32	82
Affect	5	5	10
Structure	9	9	18
Themes/Ideas	15	11	26
Tone	11	8	19
Syntax	2	0	2
Vocabulary	9	1	10
Task	102	106	208
Elements/Features	3	42	45
Cognitive Processing Behaviors	40	19	59
Physical Processing Behaviors	45	36	81
Task Experience	14	9	23
Test	38	42	80
General	23	33	56
Specific	15	9	24
Total	289	320	

Research Question 3b: Deep Dive into the Profiles

The interview data were analyzed for each profile group to garner further insights into the characteristics of the students populating the reader and learner profiles. The results from the quantitative and qualitative analyses were combined and are presented in this section.

Reader Profiles

Table 17 presents an integrated view of the results from the qualitative and quantitative analyses for the three reader profiles.

Distracted Surfers. This profile constituted the largest grouping of students (n = 68). They were the least experienced with reading evidenced by their lowest self-reported weekly reading hours (about 13 hours) and low number of books and eBooks in their possession (approximately 10 in total). The low reading experience was borne out by their learner-related metacognitive utterances (see Table 17). One of the interviewees expressed using their cursor to highlight and follow the text, which they described as a digital equivalent of fingerpoint reading, a physical behavior that is associated with emergent readers (Ehri & Sweet, 1991). Specifically, they stated: "So when I uh read it in a book, I usually use my fingers to um go through like each um words and sentences while I am reading. And that is basically like highlighting in in the computer."

The two students interviewed from the Distracted Surfer profile verbalized opposing ideas. One discussed finding reading print more challenging than listening to audio because while reading they needed to exert extra cognitive effort, and consequently could not "write down things" they deemed "important." This contrasted with the other student who found reading easier as it was more active compared to audio, which was

described as a passive experience—"doing something myself was a lot different experience than having something done read to me." The first student reported Korean as their native language, and the second student was a native English speaker.

There were also points of similarity between the two interviewees. Consistent with their low performance scores on the comprehension measures, they barely uttered anything related to the text. Only one of them mentioned not understanding "a word"—a surface-level feature. They mentioned being "distracted" and trying to get through the reading "quickly." This corroborates with the significantly lower time spent on reading the text compared to students in the other profile groups. Specifically, they spent 45% lesser time reading the experimental text than the students from the Labored Harvester group and 32% lesser than the students of the Fluent Surveyor group. This aligns with both their distraction related utterances in the interview data, and the reported poor reading habits on the quantitative measure.

Labored Harvesters. In the interviews this group professed a preference for reading on screens, considered themselves to be quick readers, and showed metacognitive awareness of the strategies they use to read. Selected representative utterances from the interviews are presented in Table 17 alongside the quantitative results. These students reported reading approximately 13 hours weekly similar to those in the Distracted Surfer profile. However, the interviewees from the Labored Harvester group were the only ones who articulated a preference for reading on screens. This screen preference, which was not captured in the quantitative data could explain the unique physical text processing behaviors associated with this group of students.

The Labored Harvester students appear to rely on physical reading strategies more than other profiles' students. One of the participants explained that they do not read "linearly," while another described how they "jump to the end, and then I'll like kind of like work my way backwards." This corroborates with the fact that students in this profile scrolled forward most frequently. I also observed higher tendency to scroll backwards in the videos of students from this profile. The verbalizations from the interviews explained the quantitative scrolling data. The students in this profile appear to be using strategies to read for specific details (keyword spotting). Indeed, non-linear digital reading patterns such as the F-pattern and the spotted pattern (looking for details) have been reported in other studies of digital reading (Delgado et al., 2018; Singer Trakhman et al., 2018; Wolf, 2018).

The digital reading behavioral strategies described by the students from the Labored Harvester profile group resulted in them scoring the highest on the recall comprehension measure among the three profiles. Given that these students verbalized reading for specific details, it is unsurprising that they performed the best at remembering facts encountered directly in the text. It is possible that the students populating the Labored Harvester profile were "harvesting" the text for key information that would help them perform well on a test. This could be a strategy that they have developed and perfected for reading texts for academic purposes. If this contention is true, the students in the Labored Harvester profile would be what Alexander (2018a) calls "information managers."

Information managers are adept at ascertaining task requirements and responding diligently to get the task done but they typically do not engage in critical or reflective

analysis (Alexander, 2018a). In keeping with the "information management" orientation, the students from the Labored Harvester group spent the longest time on the comprehension measures. They expended effort to work out the superficial structure of the text but did not make intratextual connections. One of the students determined that the presented experimental text was part of a book, and therefore had "redundant" information. They also highlighted the challenging vocabulary encountered in the text.

Their text-related utterances had a major focus on vocabulary, another indication of the superficial engagement with the text. They appeared to emphasize vocabulary as a key component for drawing meaning instead of the presented ideas and arguments (see Table 17). Even in their test-related utterances, they voiced that they had knowledge of most of the words they encountered. When discussing the text, they mentioned specific details such as "Aristotle," but did not comment on the content related to Aristotle that was part of the text. This stood in contrast to the students of the Fluent Surveyor profile who talked about the themes (e.g., happiness and its multiple layers) and evaluated the message (e.g., "existential look" at happiness).

Fluent Surveyors. This profile group constituted the least number of students at $n = 14$. They spent twice as much time reading, on average, compared to the other two groups. Students in the Fluent Surveyor profile reported possessing approximately 120 books and eBooks, which was four times the number reported by those in the Labored Harvester profile, and 12 times more than the books owned by students from the Distracted Surfer group. Keeping with their reading habits, in the interviews they verbalized being "focused" and "engaged" in the reading.

In the interviews, the Fluent Surveyors expressed strong affective responses such as feeling "stupid" for a time period extending beyond the task as a reaction to realizing an error made on the comprehension measure. This shows that they continued to think about their responses beyond what was dictated by the task requirement. This level of prolonged engagement is noteworthy considering that several interviewees from other profiles expressed that the task felt like something they were doing for "class."

The Fluent Surveyors who expressed feeling stupid (negative self-judgment), turned it into a positive adaptive cognitive response by reflecting on the causes for why they made a particular error (Buchanan et al., 2016; Zimmerman & Moylan, 2009). They reasoned that they had mistaken "egalitarianism" for "elitism." Further, they reasoned that since most "-isms" have a negative connotation, the error was a case of false analogical thinking (Alexander et al., 2016). This case of error analysis is also reflective of monitoring and evaluating, which are key to self-regulation (Zimmerman & Moylan, 2009).

Also, the students from this profile expressed feeling "confused" with the text structure, which could be an indication of more active online processing (McNamara, 2001; McDaniel & Donnelley, 1996; McNamara & Shapiro, 2005). This contrasts with the positive affective response expressed by students from the Labored Harvester profile whose text-related utterances had a positive valence (e.g., "fascinating," "interested"). Their strategic response to the complex structure of the text seems to have been to spend more time dwelling on the text as captured by the longer time they held the screen stationary compared to those in other groups. The students interviewed from the Fluent Surveyor profile did not talk about other processing behaviors, be it physical, such as

scrolling, or cognitive like paraphrasing. The fact they kept the screen stationary longer than those in other groups might suggest that they were reading closely as opposed to skimming.

If the students from the Labored Harvester profile reflected an "information management" orientation, the Fluent Surveyor profile's students were most closely aligned to "knowledge building" (Alexander, 2018a). The students who were Fluent Surveyors not only critically reflected on the text, but they continued to think about it for a prolonged period. Despite the difference in the quality of their engagement, there was no statistically significant difference in comprehension performance between the students from the Labored Harvester and Fluent Surveyor group. In terms of raw scores, the students from the Fluent Surveyor profile scored the highest on main idea comprehension item. Most likely due to the small sample size of this profile, no difference was detected.

The similarity in comprehension scores between those in the Fluent Surveyor group and the Labored Harvester profile is not entirely unexpected particularly if the "knowledge building" and "information management" orientations are to be accepted. Despite critical reflection and monitoring, knowledge builders' approach is not "test preparation" (Alexander, 2018a). Indeed, one of those interviewed verbalized that they were caught off-guard when they saw the comprehension test—"realized that there was gonna be questions that I was like, Okay. Now I need to like, remember this stuff." In contrast, information managers approach the task with the explicit goal of deriving information from the text, which could be on the test. This difference in approach might account for the similarity in performance in spite of the differences in reflective engagement with the text.

The difference between the approaches adopted by the students from the Labored Harvester and Fluent Surveyor profiles can be summed up best with the aid of a metaphor. Those populating the Fluent Surveyor group were surveying the forest while those from the Labored Harvester profile were harvesting the trees.

Listener Profiles

Table 18 presents the quantitative and qualitative data pertaining to the three listener profiles as a joint display.

Inconsistently Attentive. Based on the quantitative data, the students from the Inconsistently Attentive profile had the lowest self-efficacy scores, spent the least time listening weekly, and possessed the fewest audiobooks. This quantitative data led me to conclude that this group had the least audio experience. However, the qualitative interview data revealed that they talked about their experience with "videos" and likened it to the audio experience. They verbalized that they increase the playback speed of instructional videos, a strategy that they employed in this task as well.

Those populating the Inconsistently Attentive profile increased the playback speed more than the students from the other two groups. In the interviews, they explained that adjusting audio speed was a mechanism to regulate attention and a common behavior when they view video lectures (see Table 18). This group of students appeared to be acutely aware of the need to manage attention when listening to text through the audio medium. When viewed alongside the interview data, their quantitatively captured high fidgeting behavior certainly appears to be a mechanism to regulate attention (Langner & Eickhoff, 2013; Robertson & Garavan, 2004).

Table 17

Integrated Results Matrix for Reader Profiles

Quantitative Profiles	Qualitative Interviews	Example Quotes
Distracted Surfer (DS)		
Learner Read/week: Similar to LH Possess books: Fewest Possess eBooks: Fewest	Metacognition	"I lose attention, or I have a low attention span when uh reading a long article." "I was like definitely less distracted than the audio recording."
Text	—	—
Task Duration read: Shortest Forward scroll: Medium Stationary: Least	Cognitive Processing; Physical processing; Experience	"Try to uh find like the definition or so." "Read as soon as possible, so that I can uh try to get through quickly." "Having doing something myself was a lot different experience than having something done read to me."
Test Duration recall: Medium Duration inference: Medium Duration constructed: Medium	General, Specific	"I didn't really fully comprehend what I was reading." "I didn't understand a word."
Labored Harvester (LH)		
Learner Read/week: Similar to DS Possess books: Medium Possess eBooks: Medium	Affect; Habits and Preferences; Metacognition; Self-	"I didn't feel any stress" "I read a lot of like books in general on like my phone"

	Efficacy; Topic or Prior Knowledge	"I would try to skim fast, and then I would just miss the material."
		"I do think I am a pretty fast reader."
		"My major is not relevant to this field."
Text	Affect; Structure; Syntax; Themes/Ideas; Tone; Vocabulary	"Find it like fascinating."
		"It might have been part of a book, because of how like, if redundant."
		"It has a lot of really long sentence."
		"After the mention of Aristotle they mentioned a couple of other more names."
		"They didn't write this for the main main public audience."
		"There were some words in there that were definitely above the normal or vocabulary of some people."
Task Duration read: Longest Forward scroll: Most Stationary: Medium	Elements/ Features; Cognitive Processing; Physical processing; Experience	"But I wasn't given the highlighting feature."
		"I was like again trying to find a theme."
		"I know when I like read it. It's not very like linearly."
		"This wasn't something that I wanted to be doing."

Test Duration recall: Longest Duration inference: Longest Duration constructed: Longest	General; Specific	"I knew most of the words in there." "You cannot even know, what's the meaning of this sentence" "even though you understand every word"
Fluent Surveyor (FS) Learner Read/week: Most Possess books: Most Possess eBooks: Most	Affect; Metacognition	"I felt so stupid." "I definitely look focused."
Text	Affect; Structure; Themes/Ideas	"The reading in general made me feel very confused" "He was going like up and down in different areas of what he was talking about" "Here they [western media] make all the isms bad, any kind of communism, Marxism"
Task Duration read: Least Forward scroll: Least Stationary: Most	Experience	"It was a very neutral feeling of just focused, engaged, at least when I was doing the reading"
Test Duration recall: Shortest Duration inference: Shortest Duration constructed: Shortest	General; Specific	"I have to like, retain it." "After the fact I confused egalitarianism for elitism."

Note. The first column gives quantitative results for the indicator variables in comparative terms for the three profiles

Because they were inclined to increase audio speed, they were also the fastest among the three profiles in listening to the experimental audiobook despite going back occasionally to re-listen. This re-listening strategy could also be related to their experience with video lectures. One of the interviewees also expressed that audio allows them to pause and take notes more effectively than print. It would be worthwhile to explore the relation of this behavior to the recent move to online classes necessitated by the COVID-19 lockdown.

In the interviews, one of the students verbalized a preference for audio while again referencing their video watching habits. It is worth noting that the participant in question was not a native English speaker. The same participant stated that they understood more in the audio condition. These utterances related to performance judgment were congruent with their actual performance on the inference and vocabulary gain comprehension measure (0.78 in audio vs. 0.45 in print; 0 in audio vs. -0.15 in print) but not on the recall portion (0.67 in audio vs. 0.84 in print). They scored similarly on the main idea item. Indeed, my systematic review of the literature comparing print and audio on comprehension outcomes found that audio could facilitate comprehension for EFL and ELLs (Singh & Alexander, 2022).

The students belonging to the Inconsistently Attentive profile pinpointed to a lack of domain-specific knowledge as a likely cause for their poor comprehension. This contrasted with students from the other profile groups who vocalized that the audio medium created difficulties for comprehension. In conclusion, it seems that these students belong to the profile that has adapted to and developed strategies to effectively learn from text presented auditorily.

Inattentive. Those populating the Inattentive profile were in the middle of the two groups on most quantitative indicator variables, with a few exceptions (they fidgeted the least and very rarely went back in the audio clip to re-listen). However, they constituted the lowest performing group on the main idea comprehension measure. Concomitant quantitative and qualitative data analysis helped elaborate and accounted for their low comprehension performance.

First, despite being in the middle of the three groups on quantitative measure of listening self-efficacy, the qualitative data revealed that these students verbalized a certain disinclination toward audio because of either not seeing themselves as a "podcast person" or viewing audio as an aid for struggling readers (see Table 18).

They also articulated that they found the narrator's voice was "not fluid," and they felt bored. Relatedly, the students populating this profile vocalized that listening to audio by nature, is a passive activity, and this could also explain why they were the group that observably drank and ate (off-task behaviors) the most during the audio experimental condition. Repeatedly, they verbalized being bored or disliking the audio condition.

In terms of the text, they seemed to have remained at the surface structure level of the text and get derailed by irrelevant details. The students referred to their inability to understand the vocabulary. Further, one of the students listed the ideas presented in the text but was unable to draw intratextual connections among them. This failure explains why these students were the lowest performers on the main idea comprehension item. Another interviewee talked about trying to decipher the narrator's nationality—a factor irrelevant to comprehending the text. It is possible that some students conflated the narrator for the author.

The task-related utterances and the behaviors captured quantitatively corroborated to show that students populating this profile expended little effort on maintaining attention. They were the least fidgety of the students belonging to the three groups and barely went back to re-listen. They were not actively engaged in listening and therefore, their listening duration was very close to the audio clip's running time. These students were the most passive of the three student groups.

Persistently Attentive. The students belonging to the Persistently Attentive group had the highest scores on the main idea and vocabulary gain comprehension measures. On the quantitative learner-related variables, they reported the highest listening self-efficacy, hours spent listening to audiobooks and podcasts, and number of audiobooks. Surprisingly, in the qualitative interviews, these students also verbalized just as the students from the Inattentive profile that they did not like audio (see Table 18). The key difference was that they seemed more aware of the different types of audio formats. They recognized the experimental text was an audiobook and expressed their dislike for "audible." The students from the Inconsistently Attentive profile had put videos and audio on the same level, students of the Inattentive group had conflated podcasts and audiobooks, but those belonging to the Persistently Attentive profile were discerning of the differences among the different audio formats.

Other points of broad similarities and fine distinctions were found in the qualitative interview data of the students populating the Inattentive and Persistently Attentive profiles. One of the interviewed students from the Persistently Attentive profile, too, commented that they found audio to be a more passive activity similar to what the students from the Inattentive profile had verbalized. But the students from the

Persistently Attentive profile appeared to be driven by their top-down goal to extract meaning from the text rather than by bottom-up features of the audio and were able to estimate the attention requirements of the task (Gopher & Donchin, 1986; Kahneman, 1973).

This difference in attention regulation as a consequence of their goals was evident in their gazing patterns, which included more persistent sideway gazes. Sideway gazing is most likely a physical manifestation of attention to the audio text. However, research has not investigated eye-movements when listening to audiobooks in naturalistic settings (Fischer & Zwaan, 2008; Spivey & Geng, 2001). These students also looked on-screen the most, perhaps to determine where they are in the text by noting the playback time. This behavior also indicates monitoring, a critical component of self-regulation (Zimmerman & Moylan, 2009). Further, students from this profile group went back to re-listen to the audio most frequently, which is also a clear sign of metacognitive monitoring.

Another key difference between the students from the profile groups related to the levels of depth of text processing and critical engagement with the text. While one of the students interviewed from the Inattentive group was expending cognitive resources on analyzing the positionality of the narrator, the Persistently Attentive profile's student verbalized that the author's tone and discourse appeared to be of a "baby boomer." When analyzing the content, the student from the Persistently Attentive profile pointed out that the author was comparing "two very different things," which is evidence that they adopted a critical stance and identified a faulty analogy, thereby exhibiting evidence of relational reasoning (Alexander et al., 2016).

Table 18

Integrated Results Matrix for Listener Profiles

Quantitative Profiles	Qualitative Interviews	Example Quote
Inconsistently Attentive		
Learner Self-Efficacy: Least Listen/week: Least Possess audio: Fewest	Affect; Habits and Preferences; Metacognition; Topic or Prior Knowledge	"I was no stressful." "I like to watch videos, so maybe I like to audiobook instead of reading." "So basically 1.25 usually like I increase it [playback speed] a little bit to like grab my attention a little bit more." "It's not related to the field I study."
Text	Vocabulary	"When a sentence has a lot of that kind of words."
Task Duration listen: Shortest Gaze down: Most Gaze side: Least Gaze on-screen: Least Fidget: Most Increase playback speed: Most Rewind audio: Middle	Elements/Features; Cognitive Processing; Physical Processing	"He was talking very slowly." "I find that there are some similar content." "It's just because he speaks so slow. So I make faster."
Test Duration recall: Shortest Duration inference: Shortest Duration constructed: Shortest	General	"I understand more when I listen to this audiobook."

Inattentive		
Learner Self-Efficacy: Medium Listen/week: Medium Possess audio: Medium	Affect; Habits and Preferences; Metacognition; Self-efficacy	"I have anxiety. So I play with my hair a lot." "But I'm not like a big podcast person." "it didn't feel like as much of a pressure to like really like, engage like as like if I'm reading out loud, or if I'm like sitting there like reading it silently." "I feel like when I was growing up it was always offered as like a like. If you can't read or like, if you're having trouble with the reading, then do this [audio]"
Text	Affect; Structure; Themes/Ideas; Tone	"I know it was my least favorite section of all of them." "They are going off a tangent" "You talk about happiness, and it talked about reasons of life, and then life, expectations, and all these other topics that didn't necessarily connect to one another." "I was like trying not to laugh, because it was kind of funny."
Task Duration listen: Medium Gaze down: Medium Gaze side: Medium Gaze on-screen: Medium Fidget: Least Speed fast: Medium Rewind audio: Least	Elements/Features; Cognitive Processing; Physical Processing; Experience	"It isn't it doesn't sound fluid or natural flowing." "I was trying to figure out what nationality is the speaker from." "Listen closely to most of the uh details." "It was kind of boring like sitting there listening."

Test Duration recall: Medium Duration inference: Medium Duration constructed: Medium	General; Specific	"He was saying a lot of things I just didn't know." "I had no idea about that word."
Persistently Attentive Learner Self-Efficacy: Most Listen/week: Most Possess audio: Most	Affect; Habits and Preferences; Metacognition	"I'm already like annoyed" "I know a lot of people like like audible, or whatever. I don't like that." "I was less interested, less focused, especially partly because it was an audio."
Text	Tone	"It's probably I don't know boomer or baby boom, or maybe complaining about this generation."
Task Duration listen: Longest Gaze down: Least Gaze side: Most Gaze on-screen: Most Fidget: Middle Speed fast: None Rewind audio: Most	Elements/ Features; Cognitive Processing; Physical Processing	"Why, he was so monotone with the way he was talking." "But you're comparing two very different things." "I play with my cat more willingly, because, you know, I was like wherever I'm listening."
Test Duration recall: Longest Duration inference: Longest Duration constructed: Longest	General	"so made it [the audio] harder to absorb."

Note. The first column gives quantitative results for the indicator variables in comparative terms for the three profiles.

CHAPTER V: CONCLUSIONS AND IMPLICATIONS

In this dissertation research, I sought to investigate text comprehension differences across the print and audio mediums by acknowledging and embracing the complexity of the construct of comprehension. The organizing heuristic for this person-centered, mixed methods study was Jenkins' tetrahedral model that structurally captures the interrelations among learners, texts, task, and test in influencing comprehension. The study consisted of quantitative components that led to comparisons between the two mediums and helped identify profile groups based on variables related to learner characteristics, the task context, and the performance tests (actual and perceived) for both the print and purposefully drawn from the identified profile groups. The cued retrospective interviews were coded along Jenkins' four dimensions to describe further and elaborate on the nature of the students who formed the profile groups.

In this section, I briefly describe key findings and situate them in the broader literature. Next, I acknowledge the limitations of the current investigation. Finally, I discuss the implications of this research study and identify directions for future research.

Key Findings

Print has an advantage over audio for understanding the main idea and for incidental vocabulary learning

The results from the variable-centered analysis in this study demonstrated that text comprehension is similar across print and audio on selected-response items targeting recall and inference. Print is related with better performance on constructed-response items that included main idea and vocabulary questions. These results may appear surprising when considered alongside prior medium research on comprehension

differences, which suggests that reading on paper or digitally results in similar performance on gist level questions and differences but differences on the level of details (Singer & Alexander, 2017b; Singer Trakhman et al., 2018).

There are two possible explanations for the dissonance in findings that relate to the text and the medium. First, the texts used in prior investigations were simpler in terms of the topic and the style than the texts in this study. For instance, Singer Trakhman and colleagues (2018) used excerpts on animal and plant life from a textbook. Whereas in this dissertation study, texts on happiness and inequality were taken from Steven Pinker's *Enlightenment Now.* The abstractness of the topic and the complexity of the text structure meant that participants had to attend to and integrate various arguments and supporting details to arrive at a mental model of the text and report the main idea. This could have influenced the findings of this study.

Second, the mediums compared in the studies were different. In the Singer Trakhman collection of studies, comprehension was compared across text presented on paper with digital. On the other hand, in my dissertation research the texts were presented digitally (labelled as print) or auditorily (referred to as audio). It is possible that we would have observed the same pattern of print superiority over audio for recall of key and relevant details that Singer Trakhman et al. report if I had compared paper to audio. This contention has an elegant appeal and studies comparing digital reading and audio listening have found no difference between digital and audio, whereas those comparing paper and audio report an advantage for paper across text genres and complexity levels (Furnham et al., 1990; Rogowsky et al, 2015; Stepien-Bernabe et al., 2019; Varao Sousa, 2013).

Rogowsky and colleagues (2015) conducted a study with college-educated adults who listened to or digitally read a mixed genre. Participants in the audio condition performed as well on the multiple-choice recall and inference test as those who read the text instead. Their findings match the results of my study. But when reading on paper was compared to listening to audio, the pattern of results changed. For example, Furnham et al. (1990) compared college students' performance on free and cued recall measures across paper and audio. The researchers also manipulated text complexity and used both an easy and a difficult extract from a book on insect biology. They found that the print group outperformed the audio group regardless of text complexity in both free and cued recall measures.

Reader profile groups differed on digital reading strategies and depth of text processing

The integration of quantitative and qualitative results highlighted that the students in the three reader profiles differed in their approaches to text comprehension. The students who populated the Fluent Surveyor group were more avid readers and adopted a coherence-oriented and knowledge building approach to text comprehension (Alexander, 2018a; Kintsch & van Dijk, 1978; Rapp et al., 2007). They showed evidence of more active text processing and monitoring when they described making connections to their background knowledge on "Marxism" and articulated being "*confused*" with the text. On the other hand, the students from the Labored Harvester profile, who scrolled forward most often, focused on surface-level text features (e.g., vocabulary) and were expert information managers. Their involvement with the text content was articulated in the trite

phase—"it was *interesting*." When talking about the text, they mentioned the name of the philosopher "Aristotle" but did not elaborate on the content.

The third and lowest performing student profile was the Distracted Surfers . They included those students who do not read often, and this lack of reading experience was evident in their hasty and superficial approach to the text. They were the quickest to read and their interviews included no text-related utterances and very few instances of cognitive processing behaviors.

Based on the quality of the text processing differences across the three profiles, I had expected a stark difference in comprehension performance. However, there were only significant differences in comprehension performance on the recall and main idea items. Further, the difference in recall was observed between the students of the Labored Harvester and Distracted Surfer profiles. Likewise, the performance difference on the main idea item was between the students belonging to the Fluent Surveyor and Distracted Surfer profiles. There were no differences between the performance of students who populated the Labored Harvester and Fluent Surveyor profiles. And the topic knowledge covariate only predicted membership to the Labored Harvester profile compared to membership to the Distracted Surfer profile.

This lack of significance in outcome could be an attribute of the sample size, particularly the small size of the Fluent Surveyor profile. It is also possible that the comprehension measures were not sensitive enough to capture the differences. Finally, the information management approach adopted by those who made up the Labored Harvester profile could have proved effective for the comprehension measures of this

study, which were given immediately after the text was read. It is possible that I would have observed more of a difference if the comprehension measures were delayed.

Based on their text processing behaviors, the students in the Distracted Surfer group were most similar to the "Glider" profile that Singer Trakhman and colleagues identified. Students who belonged to the Glider profile were quickest at going over the text and remained at the surface-level (Dinsmore & Alexander, 2016; Singer Trakhman et al., 2018). Based on the data from the interviews in the current study, it seems that the Distracted Surfer group included students who were quick at reading because they found the act of reading texts challenging and were therefore, unable to expend the required effort.

The students populating the Labored Harvester profile were most similar to the Sampler reader group that Singer Trakkhman et al. (2018) uncovered. They adopted deeper processing strategies but only for selected portions of the text. In fact, these students from the Labored Harvester profile discussed a keyword spotting reading strategy, wherein they scanned the text for seemingly important words and read carefully around those words.

The students in the Fluent Surveyor profile were closest to the Singer and colleagues' (2018) Regulator profile students. They were purposeful in their reading and displayed more metacognitive monitoring (Baker & Brown, 1984). Given the complexity of the text, which included several arguments and supporting details that needed to be integrated to form a coherent text model, the students who populated the Fluent Surveyor profile did not scroll as much as the other profiles and held the screen stationary to dwell on the text.

Research shows that students' understanding of the text (and not recall) benefits significantly from texts that require effortful processing (Kintsch & Young, 1984; McDaniel & Donnelley, 1996; Wiley & Voss, 1999). But comprehension is situated not only in the text. Students will benefit from a complex text only under the condition that the *learner is willing* to expend the effort to process it. Based on their high performance on the main idea item in terms of the raw score, this group was arguably the most willing to expend the cognitive effort to critically engage with the text excerpt from Steven Pinker's *Enlightenment Now.*

Listener profile groups are distinguishable on attention regulation, preferences, and depth of text processing

This mixed methods investigation was the first exploratory investigation into the nature of listener profiles. I uncovered three meaningful groups based on indicator variables. The profiles were verified on comprehension and calibration outcomes. The quantitative analysis revealed that the most salient factor differentiating the three profiles was their off-task behaviors, including fidgeting and gazing patterns. These behaviors were most likely a manifestation of attention regulation strategies.

The qualitative data corroborated that attention was a recurring theme voiced by interviewees across profiles. The factor that set the profile groups apart was *how* the learners chose to regulate attention. It appears that those part of the Inconsistently Attentive profile relied on non-disruptive, continuous fidgeting behaviors (e.g., playing with their hair) to maintain attention. Students who constituted the Inattentive profile passively listened to the audio, while the students from the Persistently Attentive group

looked sideways and monitored the time remaining by glancing back at the screen, an indication of self-regulation (Zimmerman & Moylan, 2009).

Unsurprisingly, students who formed the Persistently Attentive profile, spent the longest listening to the text and re-listened the most were the highest performers on main idea and vocabulary gain (the two comprehension dimensions where I found a significant effect). The students populating the Inattentive group, were faster than the students of the Persistently Attentive profile but slower than students from the Inconsistently Attentive, were at the lowest level in terms of performance.

Students from the Inconsistently Attentive group were the fastest as they increased the playback speed the most, a strategy reportedly used to regulate attention. In the interviews, one of these students explained that they increased audio speed gradually to get acclimated and used the speed changes to maintain attention. They had comparable performance as the students populating the Persistently Attentive profile on main idea (gist understanding). But they had poorer performance than the students from the Persistently Attentive group on vocabulary gain. It seems that the Inconsistently Attentive profile's students had the most listening strategies. They were aware of the need to regulate attention and purposefully managed their attention through strategies such as increasing speed or behaviors like fidgeting.

Unexpectedly, the students populating the Inconsistently Attentive group reported spending the least time listening to audio weekly and possessing the fewest audiobook on the quantitative self-report measure. But through the qualitative part of my study, I learned that these students watched videos and expressed their preference for the video medium to learn content. It is possible that video-watching strategies effectively transfer

to the audio medium, given the similarities in affordances. They also wondered if they preferred audio precisely because they liked videos. On the other hand, one of the interviewees from the Inattentive profile voiced a strong dislike for video lectures in the interviews and told me that if my study aimed to determine if online classes work, I should let it be known that they do not. Another interviewee from the same profile also expressed that the study triggered unpleasant memories of online classes.

The final aspect where the three profiles differed in an expected manner was the depth of text processing reflected in their comprehension performance and data from the interviews. The students populating the Inconsistently Attentive profile, who actively managed their attention talked about the similarity of content in the reading and listening condition, evidence that they were making intertextual connections (Alexander & the DRLRL, 2020; Kendeou et al., 2017). The students who formed the Inattentive group, found it challenging to maintain attention and disliked audio, talked about trying to determine the narrator's nationality, showing evidence of getting derailed by seductive details (Alexander, 2019; Garner et al., 1991). The students from the Persistently Attentive profile, who were regulated and effortful, talked about the author's positionality, showing evidence of higher-order thinking (Afflerbach et al., 2015; Afflerbach et al., 2008).

Limitations

The limitations of this investigation should be kept in mind when interpreting the results, methodological contributions, and practical implications. First, the study was conducted with undergraduate students in the United States who represent a population of mature readers with a high level of experience with reading digital print texts (Alexander,

2005; Huang et al., 2014). Therefore, the generalizability of the findings to other learner groups is suspect. Further, I was unable to meet my proposed sample size, which may have changed the nature of the profiles.

Second, I included a number of variables related to the learner but at the expense of excluding some key cognitive and affective variables related to text comprehension. In the context of this study particularly, measures of cognitive variables such as attention and working memory would have compounded our understanding of comprehension differences within the audio medium and across the two mediums under scrutiny (Kendeou et al., 2014; Siegel, 1994). In fact, several interviewees alluded to the difficulty with attention and the problem with only having their "brains to hold information" in the audio condition. Affective factors that research has established as critical that were not part of this study but would have enriched or even modified the nature of the profiles were interest, motivation, and socio-cultural background (Bauer et al., 2021; Wigfield et al., 2008).

Third, the data collection methods could have biased the results. The processing data for the quantitative portion of the study were extracted through manual coding of Zoom® screen share recordings. Participants could have altered their behaviors because they knew I would be watching the videos, resulting in the "Hawthorne effect" (Adair, 1984; McCambridge et al., 2014). The qualitative data were gathered through cued retrospective interviews. The specific context of the interview wherein I played back video clips (participant as observer) and the passage of time could have influenced the account of the interviewees. Another factor influencing recall in the interviews could be variable levels of self-awareness (Nigro & Neisser, 1983). Also, the interviews were

conducted with a small group of students who agreed to participate in this part of the study. Consequently, the insights from the qualitative analysis were drawn from a small set of self-selected individuals.

The fourth limitation also relates to methodological considerations. I used an improvised method to capture processing data which included gazing patterns. These data were extracted through manual coding, a labor-intensive process that is sensitive to human errors. Future studies should explore more automated and reliable methods of collecting eye tracking, scrolling, and cursor movement data. Finally, the comprehension outcome variables were measured in immediate posttests. The pattern of results could be altered if I had also included a delayed posttest.

Contribution

While scientific literature is replete with theoretical and empirical accounts of text comprehension, less is known how the complex set of factors related to learners and their contexts (text, task, and test) influence comprehension across different mediums, particularly audio (Singh & Alexander, 2022). The substantive contribution of this exploratory study is that it sheds light on comprehension differences across print and audio and describes different reader and listener profiles within undergraduate students while embracing the complexity of the comprehension construct (Jenkins, 1977; 1979; McMaster & Kendeou, 2023). This is also the first study to profile listeners as they engaged with excerpts from a mixed genre audiobook. Constructing reader and listener profiles can assist in creating personalized learning solutions for students.

Methodologically, the consideration of the complexity of comprehension is evident in the conceptual foundation of the study (Jenkins' tetrahedral model), the variety

of measures (self-report measures, videos for behavioral data, cognitive measures of performance), and the overall mixed-methods approach adopted in this study. In addition to enriching the understanding of the profile groups, the qualitative part of the study was pivotal in providing insights into areas for future research (see Directions for Future Research).

This research study on print and audio expands the universe of text comprehension research to reflect the realities of students' digital experiences with text, which are no longer limited to print texts. In keeping with this investigation's rootedness in reality, its major contribution is in the practical arena.

Implications for Practice

This study has several implications for educators and instructional designers hoping to support students' text comprehension in digital contexts. The results from this study indicate that audiobooks can be as effective as digital books for recall and inference but not for integrating the main message. It follows that educators should keep the learning goal in mind when selecting between print or audio for content delivery. This is particularly important for online classes where there may be a higher reliance on the audio medium.

Educators should emphasize transferable comprehension competencies rather than focusing on medium-specific strategies. The results from this study indicated that inferencing ability, understanding of text structure, prioritizing information needed to derive meaning, and constructing mental models of the text were critical for successful comprehension irrespective of the medium. But the strategies reported by students did not support these abilities. In fact, those interviewed from the low performing profile groups

described taking a keyword approach to their reading, wherein they looked for specific words and phrases that they deemed relevant to the topic. They explained that this was their preferred way of reading digital texts, which explained their scrolling patterns. However, this medium-specific strategy does not work for complex texts such as the ones used in this study. Those who performed well simply took their time to read and integrate the information while critically engaging with the assertions made by the author.

Three interrelated areas for learning with audiobooks are worth highlighting because of their practical implications: (a) attention regulation; (b) narrator's voice; (c) spatial anchoring. Students' gazing patterns, off-task behaviors, and their interviews corroborated that they were unable to sustain focus and had trouble concentrating on the audio text. Some participants attributed this to the quality of the narrator's voice while others mentioned that they found it hard to focus because they could not track where they were in the text or re-listen to it. Features such as keyword search, making the rewind button more intuitive, embedding headings within the audio timeline, and including an accessible transcript could be incorporated in applications for playing audio content, especially in those designed for instructional purposes.

Directions for Future Research

There are several directions for future research that would extend and build on the findings of this investigation of text comprehension across different mediums that I enumerate here. I contend that comparing print and audio is an area of research that is critical given the rising use of podcasts, audiobooks, and video lectures. Based on the findings from this study

First, research is needed to account for the variable experiences, behaviors, and outcomes of learners with different language backgrounds and age groups. In the qualitative interviews, a non-native English speaker voiced that they preferred audio over print for learning. This insightful comment combined with the positive impact of audio on comprehension for English language learners, highlights the need to explore how the students' language status influences text comprehension across different mediums. Further research is also required to determine the developmental trajectories of learning with text presented through different mediums.

Second, unpacking how students' goals influence learning from text across mediums is an important line of research. Several interviewees expressed that they were not invested in the experiment and treated it as something they would do for class. It would be worthwhile to explore how learner goals and purposes (leisure vs. academic) influence comprehension across different mediums.

Third, understanding if and how the narrator's voice, tone, and gender plays a role in audio comprehension would be useful. The interview data showed that the narrator was a key figure who influenced the audio experience. Further research into how the narrators voice and identity influence engagement in audio and with the textual content would contribute to understanding the nuances of text comprehension with audio.

Fourth, experimenting with different text types and genres in future research would contribute to this research area. In this study, mixed genre texts on complex phenomena like happiness and inequality were presented. Further work needs to be done with different text genres (e.g., narrative, exposition), texts about varied topics (e.g., scientific, literary), and purposes (e.g., informational, refutational).

Fifth, conducting more ecologically valid research would be a productive contribution to this field. It is essential to capture how individuals listen to audio in their everyday lives. Current studies, including this one, does not capture how individual listen to audio, be it podcasts or audiobooks. Scientists could study more naturalistic contexts of audio usage such as, listening while walking, doing chores, or exercising.

APPENDICES

APPENDIX A: Experimental Texts

Segment 1: Inequality

The confusion of inequality with poverty comes straight out of the lump fallacy—the mindset in which wealth is a finite resource, like an antelope carcass, which has to be divvied up in zero-sum fashion, so that if some people end up with more, others must have less. As we just saw, wealth is not like that: since the Industrial Revolution, it has expanded exponentially. That means that when the rich get richer, the poor can get richer, too. Even experts repeat the lump fallacy, presumably out of rhetorical zeal rather than conceptual confusion. Thomas Piketty, whose 2014 bestseller Capital in the Twenty-First Century became a talisman in the uproar over inequality, wrote, "The poorer half of the population are as poor today as they were in the past, with barely 5 percent of total wealth in 2010, just as in 1910." But total wealth today is vastly greater than it was in 1910, so if the poorer half own the same proportion, they are far richer, not "as poor."

A more damaging consequence of the lump fallacy is the belief that if some people get richer, they must have stolen more than their share from everyone else. A famous illustration by the philosopher Robert Nozick, updated for the 21st century, shows why this is wrong. Among the world's billionaires is J. K. Rowling, author of the Harry Potter novels, which have sold more than 400 million copies and have been adapted into a series of films seen by a similar number of people. Suppose that a billion people have handed over $10 each for the pleasure of a Harry Potter paperback or movie ticket, with a tenth of the proceeds going to Rowling. She has become a billionaire, increasing inequality, but she has made people better off, not worse off (which is not to say that every rich person has made people better off). This doesn't mean that Rowling's wealth is just deserts for her effort or skill, or a reward for the literacy and happiness she added to the world; no committee ever judged that she deserved to be that rich. Her wealth arose as a by-product of the voluntary decisions of billions of book buyers and moviegoers.

To be sure, there may be reasons to worry about inequality itself, not just poverty. Perhaps most people are like Igor and their happiness is determined by how they compare with their fellow citizens rather than how well-off they are in absolute terms. When the rich get too rich, everyone else feels poor, so inequality lowers well-being even if everyone gets richer. This is an old idea in social psychology, variously called the theory of social comparison, reference groups, status anxiety, or relative deprivation. But the idea must be kept in perspective. Imagine Seema, an illiterate woman in a poor country who is village-bound, has lost half her children to disease, and will die at fifty, as do most of the people she knows. Now imagine Sally, an educated person in a rich country who has visited several cities and national parks, has seen her children grow up, and will live

to eighty, but is stuck in the lower middle class. It's conceivable that Sally, demoralized by the conspicuous wealth she will never attain, is not particularly happy, and she might even be unhappier than Seema, who is grateful for small mercies. Yet it would be mad to suppose that Sally is not better off, and positively depraved to conclude that one may as well not try to improve Seema's life because it might improve her neighbors' lives even more and leave her no happier.

In any case, the thought experiment is moot, because in real life Sally almost certainly is happier. Contrary to an earlier belief that people are so mindful of their richer compatriots that they keep resetting their internal happiness meter to the baseline no matter how well they are doing, we will see in chapter 18 that richer people and people in richer countries are (on average) happier than poorer people and people in poorer countries.

But even if people are happier when they and their countries get richer, might they become more miserable if others around them are still richer than they are—that is, as economic inequality increases? In their well-known book The Spirit Level, the epidemiologists Richard Wilkinson and Kate Pickett claim that countries with greater income inequality also have higher rates of homicide, imprisonment, teen pregnancy, infant mortality, physical and mental illness, social distrust, obesity, and substance abuse. The economic inequality causes the ills, they argue: unequal societies make people feel that they are pitted in a winner-take-all competition for dominance, and the stress makes them sick and self-destructive.

The Spirit Level theory has been called "the left's new theory of everything," and it is as problematic as any other theory that leaps from a tangle of correlations to a single-cause explanation. For one thing, it's not obvious that people are whipped into competitive anxiety by the existence of J. K. Rowling and Sergey Brin as opposed to their own, local rivals for professional, romantic, and social success. Worse, economically egalitarian countries like Sweden and France differ from lopsided countries like Brazil and South Africa in many ways other than their income distribution. The egalitarian countries are, among other things, richer, better educated, better governed, and more culturally homogeneous, so a raw correlation between inequality and happiness (or any other social good) may show only that there are many reasons why it's better to live in Denmark than in Uganda. Wilkinson and Pickett's sample was restricted to developed countries, but even within that sample the correlations are evanescent, coming and going with choices about which countries to include. Wealthy but unequal countries, such as Singapore and Hong Kong, are often socially healthier than poorer but more equal countries, such as those of ex-Communist Eastern Europe.

Most damagingly, the sociologists Jonathan Kelley and Mariah Evans have snipped the causal link joining inequality to happiness in a study of two hundred thousand people in sixty-eight societies over three decades. (We will examine how happiness and life satisfaction are measured in chapter 18.) Kelley and Evans held

constant the major factors that are known to affect happiness, including GDP per capita, age, sex, education, marital status, and religious attendance, and found that the theory that inequality causes unhappiness "comes to shipwreck on the rock of the facts." In developing countries, inequality is not dispiriting but heartening: people in the more unequal societies are happier. The authors suggest that whatever envy, status anxiety, or relative deprivation people may feel in poor, unequal countries is swamped by hope. Inequality is seen as a harbinger of opportunity, a sign that education and other routes to upward mobility might pay off for them and their children. Among developed countries (other than formerly Communist ones), inequality made no difference one way or another. (In formerly Communist countries, the effects were also equivocal: inequality hurt the aging generation that grew up under communism, but helped or made no difference to the younger generations.)

Segment 2: Happiness 1

If popular impressions are a guide, today's Americans are not one and a half times happier (as they would be if happiness tracked income), or a third happier (if it tracked education), or even an eighth happier (if it tracked longevity). People seem to bitch, moan, whine, carp, and kvetch as much as ever, and the proportion of Americans who tell pollsters that they are happy has remained steady for decades. Popular culture has noticed the ingratitude in the Internet meme and Twitter hashtag #first world problems and in a monologue by the comedian Louis C.K. known as "Everything's Amazing and Nobody's Happy":

When I read things like, "The foundations of capitalism are shattering," I'm like, maybe we need some time where we're walking around with a donkey with pots clanging on the sides. . . . 'Cause now we live in an amazing world, and it's wasted on the crappiest generation of spoiled idiots. . . . Flying is the worst one, because people come back from flights, and they tell you their story. . . . They're like, "It was the worst day of my life. . . . We get on the plane and they made us sit there on the runway for forty minutes." . . . Oh really, then what happened next? Did you fly through the air, incredibly, like a bird? Did you soar into the clouds, impossibly? Did you partake in the miracle of human flight, and then land softly on giant tires that you couldn't even conceive how they fuckin' put air in them? . . . You're sitting in a chair in the sky. You're like a Greek myth right now! . . . People say there's delays? . . . Air travel's too slow? New York to California in five hours. That used to take thirty years! And a bunch of you would die on the way there, and you'd get shot in the neck with an arrow, and the other passengers would just bury you and put a stick there with your hat on it and keep walking. . . . The Wright Brothers would kick us all in the [crotch] if they knew.

Writing in 1999, John Mueller summed up the common understanding of modernity at the time: "People seem simply to have taken the remarkable economic improvement in stride and have deftly found new concerns to get upset about. In an

important sense, then, things never get better." The understanding was based on more than just impressions of American malaise. In 1973 the economist Richard Easterlin identified a paradox that has since been named for him. Though in comparisons within a country richer people are happier, in comparisons across countries the richer ones appeared to be no happier than poorer ones. And in comparisons over time, people did not appear to get happier as their countries got richer.

The Easterlin paradox was explained with two theories from psychology. According to the theory of the hedonic treadmill, people adapt to changes in their fortunes, like eyes adapting to light or darkness, and quickly return to a genetically determined baseline. According to the theory of social comparison (or reference groups, status anxiety, or relative deprivation, which we examined in chapter 9), people's happiness is determined by how well they think they are doing relative to their compatriots, so as the country as a whole gets richer, no one feels happier—indeed, if their country becomes more unequal, then even if they get richer they may feel worse.

If, in this sense, things never get better, one can wonder whether all that economic, medical, and technological so-called progress was worth it. Many argue that it was not. We have been spiritually impoverished, they say, by the rise of individualism, materialism, consumerism, and decadent wealth, and by the erosion of traditional communities with their hearty social bonds and their sense of meaning and purpose bestowed by religion. That is why, one often reads, depression, anxiety, loneliness, and suicide have been soaring, and why Sweden, that secular paradise, has a famously high rate of suicide. In 2016 the activist George Monbiot prosecuted the cultural pessimist's time-honored campaign against modernity in an op-ed entitled "Neoliberalism Is Creating Loneliness. That's What's Wrenching Society Apart." The tag line was, "Epidemics of mental illness are crushing the minds and bodies of millions. It's time to ask where we are heading and why." The article itself warned, "The latest, catastrophic figures for children's mental health in England reflect a global crisis."

If all those extra years of life and health, all that additional knowledge and leisure and breadth of experience, all those advances in peace and safety and democracy and rights, have really left us no happier but just lonelier and more suicidal, it would be history's greatest joke on humanity. But before we start walking around with a donkey with pots clanging on the sides, we had better take a closer look at the facts about human happiness.

At least since the Axial Age, thinkers have deliberated about what makes for a good life, and today happiness has become a major topic in social science. Some intellectuals are incredulous, even offended, that happiness has become a subject for economists rather than just poets, essayists, and philosophers. But the approaches are not opposed. Social scientists often begin their studies of happiness with ideas that were first conceived by artists and philosophers, and they can pose questions about historical and global patterns that cannot be answered by solitary reflection, no matter how insightful.

That is especially true for the question of whether progress has left people happier. To answer it, we must first assuage the critics' incredulity over the possibility that happiness can even be measured.

Artists, philosophers, and social scientists agree that well-being is not a single dimension. People can be better off in some ways and worse off in others. Let's distinguish the major ones.

We can begin with objective aspects of well-being: the gifts we deem intrinsically worthwhile whether or not their possessors appreciate them. At the top of that list is life itself; also on it are health, education, freedom, and leisure. That is the mindset behind Louis C.K.'s social criticism and, in part, behind Amartya Sen's and Martha Nussbaum's conceptions of fundamental human capabilities. In this sense we can say that people who live long, healthy, and stimulating lives are truly better off even if they have a morose temperament or are in a bad mood or are spoiled idiots and fail to count their blessings. One rationale for this apparent paternalism is that life, health, and freedom are prerequisites to everything else, including the very act of pondering what is worthwhile in life, and so they are worthy by their very nature. Another is that the people who have the luxury of failing to appreciate their good fortune make up a biased sample of lucky survivors. If we could canvass the souls of the dead children and mothers and the victims of war and starvation and disease, or if we went back in time and gave them a choice between proceeding with their lives in a premodern or modern world, we might uncover an appreciation of modernity that is more commensurate with its objective benefits. These dimensions of well-being have been the topics of the preceding chapters, and the verdict on whether they have improved over time is in.

Segment 3: Happiness 2

In theory, freedom is independent of happiness. People can surrender to fatal attractions, crave pleasures that are bad for them, regret a choice the morning after, or ignore advice to be careful what they wish for. In practice, freedom and the other good things in life go together. Whether assessed objectively through a democracy index for a country as a whole, or subjectively through people's ratings of whether they feel they have "free choice and control over their lives," the level of happiness in a country is correlated with the level of freedom. Also, people single out freedom as a component of a meaningful life, whether or not it leads to a happy life. Like Frank Sinatra, they may have regrets, they may take blows, but they do it their way. People can even value autonomy over happiness: many who have gone through a painful divorce, for example, would still not choose to return to a time when their parents would have arranged their marriages.

What about happiness itself? How can a scientist measure something as subjective as subjective well-being? The best way to find out how happy people are is to ask them. Who could be a better judge? An old Saturday Night Live skit has Gilda Radner in a postcoital conversation with a nervous lover (played by Chevy Chase) who is worried she

didn't have an orgasm, and she consoles him by saying, "Sometimes I do and I don't even know it." We laugh because when it comes to subjective experience, the experiencer herself is the ultimate authority. But we don't have to take people's word for it: self-reports of well-being turn out to correlate with everything else we think of as indicating happiness, including smiles, a buoyant demeanor, activity in the parts of the brain that respond to cute babies, and, Gilda and Chevy notwithstanding, judgments by other people.

Happiness has two sides, an experiential or emotional side, and an evaluative or cognitive side. The experiential component consists of a balance between positive emotions like elation, joy, pride, and delight, and negative emotions like worry, anger, and sadness. Scientists can sample these experiences in real time by having people wear a beeper that goes off at random times and prompts them to indicate how they are feeling. The ultimate measure of happiness would consist of a lifetime integral or weighted sum of how happy people are feeling and how long they feel that way. Though experience sampling is the most direct way of assessing subjective well-being, it's laborious and expensive, and there are no good datasets that compare people in different countries or track them over the years. The next best thing is to ask people how they are feeling at the time, or how they remember having felt during the day or week before.

This brings us to the other side of well-being, people's evaluations of how they are living their lives. People can be asked to reflect on how satisfied they feel "these days" or "as a whole" or "taking all things together," or to render the almost philosophical judgment of where they stand on a ten-rung ladder ranging from "the worst possible life for you" to "the best possible life for you." People find these questions hard (not surprisingly, since they are hard), and their responses may be warped by the weather, their current mood, and what they were asked about immediately beforehand (with questions to college students about their dating life, or to anyone about politics, having a reliably depressive effect). Social scientists have become resigned to the fact that happiness, satisfaction, and best-versus-worst-possible life are blurred in people's minds and that it's often easiest just to average them together.

Emotions and evaluations are, of course, related, though imperfectly: an abundance of happiness makes for a better life, but an absence of worry and sadness does not. And this brings us to the final dimension of a good life, meaning and purpose. This is the quality that, together with happiness, goes into Aristotle's ideal of eudaemonia or "good spirit." Happiness isn't everything. We can make choices that leave us unhappy in the short term but fulfilled over the course of a life, such as raising a child, writing a book, or fighting for a worthy cause.

Though no mortal can stipulate what really makes a life meaningful, the psychologist Roy Baumeister and his colleagues probed for what makes people feel their lives are meaningful. The respondents separately rated how happy and how meaningful their lives were, and they answered a long list of questions about their thoughts,

activities, and circumstances. The results suggest that many of the things that make people happy also make their lives meaningful, such as being connected to others, feeling productive, and not being alone or bored. But other things can make lives happier while leaving them no more meaningful or even less so. People who lead happy but not necessarily meaningful lives have all their needs satisfied: they are healthy, have enough money, and feel good a lot of the time.

People who lead meaningful lives may enjoy none of these boons. Happy people live in the present; those with meaningful lives have a narrative about their past and a plan for the future. Those with happy but meaningless lives are takers and beneficiaries; those with meaningful but unhappy lives are givers and benefactors. Parents get meaning from their children, but not necessarily happiness. Time spent with friends makes a life happier; time spent with loved ones makes it more meaningful. Stress, worry, arguments, challenges, and struggles make a life unhappier but more meaningful. It's not that people with meaningful lives masochistically go looking for trouble but that they pursue ambitious goals: "Man plans and God laughs." Finally, meaning is about expressing rather than satisfying the self: it is enhanced by activities that define the person and build a reputation.

We can see happiness as the output of an ancient biological feedback system that tracks our progress in pursuing auspicious signs of fitness in a natural environment. We are happier, in general, when we are healthy, comfortable, safe, provisioned, socially connected, sexual, and loved. The function of happiness is to goad us into seeking the keys to fitness: when we are unhappy, we scramble for things that would improve our lot; when we are happy, we cherish the status quo. Meaning, in contrast, registers the novel and expansive goals that are opened up for us as social, brainy, and talkative occupants of the uniquely human cognitive niche. We consider goals that are rooted in the distant past and stretch far into the future, that affect people beyond our circle of acquaintance, and that must be ratified by our fellows, based on our ability to persuade them of their worth and on our reputation for benevolence and efficacy. An implication of the circumscribed role of happiness in human psychology is that the goal of progress cannot be to increase happiness indefinitely, in the hope that more and more people will become more and more euphoric. But there is plenty of unhappiness that can be reduced, and no limit as to how meaningful our lives can become.

APPENDIX B: Pre Reading and Listening

Demographic Questionnaire

Q1. Gender identification:

- o Woman
- o Man
- o Non-binary/ third gender
- o Prefer to self-describe
- o Prefer not to say

Q2. Class standing:

- o Freshman
- o Sophomore
- o Junior
- o Senior
- o Graduate student
- o Others (please specify)

Q3. Major:

Q4. Age (in years):

Q5. What is your native language or languages?

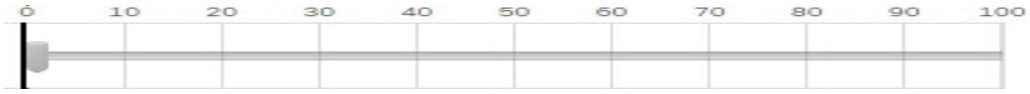
Q6. Email

Self-Efficacy Measure

Direction. Rate your confidence in being able to read/listen and understand what the author/narrator was saying in each of the following texts: (0 – not confident at all; 100 – highly confident)

c) A short fiction story.

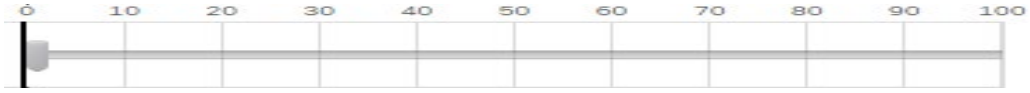
- Print



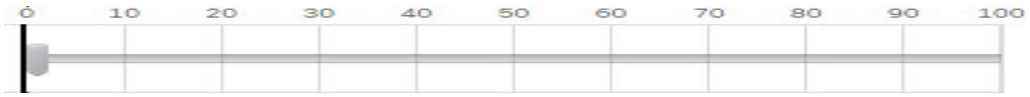
- Audio

d) A 400-page novel.

- *Print*



- *Audio*



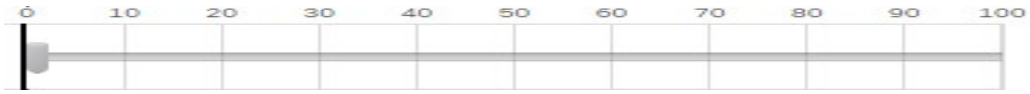
e) A textbook in your major field.

- Print

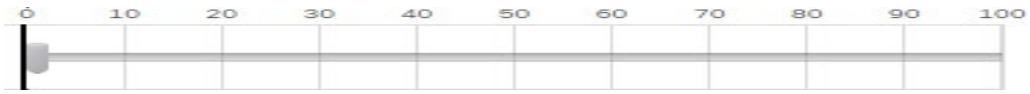
- Audio

f) A scholarly article in a journal in your field.

- Print

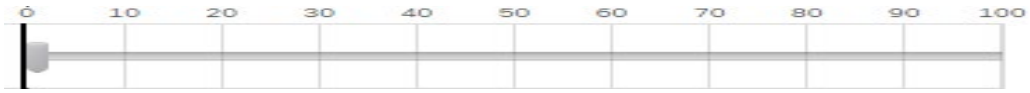


- Audio

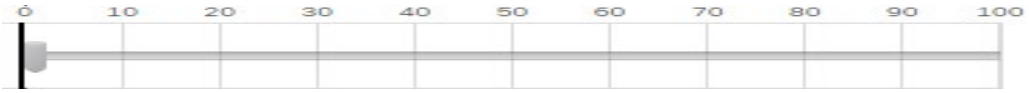


g) The daily newspaper.

- Print

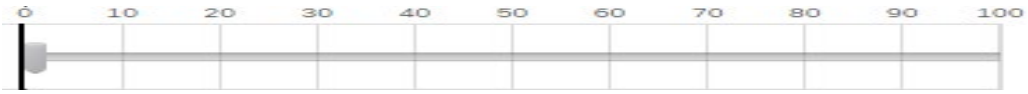


- Audio

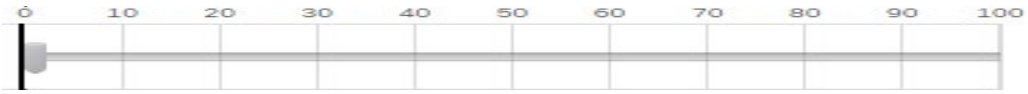


h) An article in a magazine like Time.

- Print



- Audio



i) Recognize the main points or theme in a passage.

- Print

- Audio

j) Use previous knowledge to help understand new material.

- Print

- Audio

Reading and Listening Habits Measure

In a typical week, how many hours do you spend reading the following:

Reading for leisure

- Books (e.g., novels, non-fiction, biographies)_____
- Other forms of leisure reading (e.g., news stories, magazine articles, op-eds, blog posts)

_____

Reading for school

- Textbooks

- Other forms of school reading (e.g., teacher handouts, slides, assigned readings)

  _____

  In a typical week, how many hours do you spend listening:

  Listening for leisure

- Audiobooks (e.g., novels, non-fiction, biographies)_____
- Podcasts _____

  Listening for school

- Audio textbooks
- Other forms of school listening (e.g., using text to speech feature for slides, teacher handouts) _____

  Approximately how many of the following forms of text, do you have in your possession?

- Books: ☐
- e-books: ☐
- Audiobooks: ☐

Topic Knowledge Test

Directions. Provide a definition for each of the words displayed. If you are not certain of the definition, share what you think the meaning of the word may be.

Q1. paradox

_____

Q2. materialism

_____

Q3. warped

_____

Q4. masochistic

_____

Q5. fallacy

_____

Q6. egalitarian

_____

APPENDIX C: Post Reading and Listening

Comprehension Test and Calibration of Performance

Segment 1: Inequality

Selected Response Items: Recall of Explicit Information

Q1. According to the author, one of the consequences of the lump fallacy is a belief that:
- o poor people need to work harder to improve their life
- o people have become richer by stealing from others
- o inequality and poverty are unrelated to each other
- o it is a moral imperative for everyone to have enough

Q2. What is the point that the author is trying to make by including the J.K. Rowling example?
- o Billionaires like her are one of the main causes of rising economic inequality
- o She is one of the richest people because of the Harry Potter books and films
- o Her wealth is a consequence of the spending decisions of consumers
- o She deserves to be rich because her books have made a significant contribution

Q3. What is the comparison that the author is making between France and Brazil?
- o France and Brazil differ on many aspects including income distribution
- o There is greater equality in France because it is better governed
- o The economically disadvantaged in France and Brazil face similar hardships
- o There is greater cultural diversity in France than in Brazil

Q4. According to research described in the chapter, in formerly communist countries:
- o sense of inequality is harmful for older people only
- o there are fewer job opportunities that can improve lives
- o governments are more committed to reducing inequality
- o people are happier because they live in an equal society

Selected Response Items: Inference

Q5. The author juxtaposes the views of scientists on inequality to:
- o illustrate the limitations in the current research on inequality
- o indicate that inequality must be studied alongside other factors
- o demonstrate how particular social ills arise from inequality
- o show how scientists can reach different interpretations from the same data

Q6. Which statement about the author is best supported by the information in the chapter?
- o He argues that poverty is society's real problem
- o He is an admirer of the economist Thomas Pikketty
- o He believes that inequality is the defining challenge of our time

o He believes that wealth has increased in the modern age

Q7. The author most likely refers to the Seema and Sally story to emphasize that:

o sanitary conditions are critical for improving women's health
o contentment can be realized in both rural and urban settings
o conspicuous wealth in certain countries leads to social distrust
o it is more critical to improve living conditions than reduce inequality

Constructed Response Items: Main Idea

What was the main idea of this passage? That is, what was the author trying to communicate? Justify your response.

_____

_____

_____

_____

Constructed Response Items: Vocabulary

Directions. Provide a definition for each of the words displayed. If you are not certain of the definition, share what you think the meaning of the word may be.

Q1. fallacy

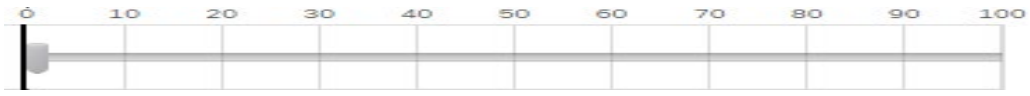
_____

Q2. egalitarian

_____

Calibration of Performance

Direction. On a scale of 0 to 100, rate your performance on the comprehension questions you just answered (0 –very poor; 100 – excellent):



Segment 2: Happiness 1

Selected Response Items: Recall of Explicit Information

Q1. What example does Louis C.K. use to make his point about people's tendency to gripe?

o Air travel
o Healthcare
o Education

o Internet

Q2. The author says that well-being entails:

o community, education, health
o life, education, leisure
o community, spirituality, satisfaction
o spirituality, life, satisfaction

Q3. The theory of the hedonic treadmill tells us that:

o the more you chase happiness the harder it is to get it
o happiness levels permanently change with high income levels
o exercising more leads to higher levels of happiness
o humans quickly adapt to changes in their fortunes

Q4. Named after an economist, the Easterlin paradox states that:

o economic growth is the primary driver of well-being
o higher income does not make people happier
o people in poorer countries are happier than people in richer countries
o countries do not pursue happiness as the goal of development

Selected Response Items: Inference

Q5. Which of the following statements best describes the author's position on well-being?

o Well-being consists of multiple components
o Well-being is not a consequence of progress
o Social comparisons prevent us from experiencing well-being
o Well-being can only be conceived by artists and philosophers

Q6. The author mentions the hashtag #firstworldproblems to highlight:

o The wealth gap between the developed and developing world
o Increasing levels of satisfaction in American society with rising income levels
o Developed countries, such as Sweden, with high suicide rates
o Failure of people to see how much happier they are compared to those in the past

Q7. Which of the following statements best describes the author's view about the position expressed in the op-ed, "Neoliberalism Is Creating Loneliness?"

o There is an epidemic of loneliness in the developed world
o Writers agree that competition is the root of society's problems
o There is no happiness without connections between human beings
o The evidences in support of decreasing happiness levels are not objective

Constructed Response Items: Main Idea

What was the main idea of this passage? That is, what was the author trying to communicate? Justify your response.

_____

_____

_____

_____

Constructed Response Items: Vocabulary

Directions. Provide a definition for each of the words displayed. If you are not certain of the definition, share what you think the meaning of the word may be.

Q1. paradox

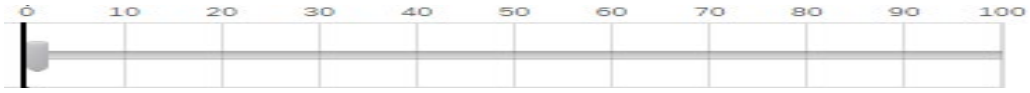
_____

Q2. materialism

_____

Calibration of Performance

Direction. On a scale of 0 to 100, rate your performance on the comprehension questions you just answered (0 –very poor; 100 – excellent):



Segment 3: Happiness 2

Selected Response Items: Recall of Explicit Information

Q1. The author says that in practice freedom and happiness:
- o are correlated to each other
- o are independent of each other
- o cannot be measured objectively
- o are one and the same thing

Q2. Happiness has two sides. They are:
- o experiential and evaluative
- o physical and emotional
- o momentary and long-term
- o personal and national

Q3. According to the author, progress is achieved by:
- o reducing unhappiness
- o increasing life expectancy
- o increasing capitalism
- o reducing choices

Q4 According to the author, what would be the ultimate way to measure happiness?
- o Making people rate how meaningful their lives are or have been
- o Measuring people's emotions over a lifetime through experience sampling

- o Tracking the rates of depression and comparing across countries
- o Asking people how they have been feeling in a particular day

Selected Response Items: Inference

Q5. Which of the following best describes the author's idea about the relation between happiness and meaning in life?

- o One cannot have happiness and meaning in life at the same time
- o The quest for meaning is pointless unless it results in happiness
- o Happy people tend to be richer than people who lead meaningful lives
- o Meaning is worth pursuing even when it does not lead to happiness

Q6. According to the author, which of the following best describes people's evaluations of happiness?

- o Evaluations of happiness are impacted by the current weather conditions
- o Levels of happiness cannot be evaluated by relying on people's emotions
- o People only take positive emotions into account when evaluating happiness
- o Absence of worry and sadness leads to better evaluations of happiness

Q7. Which statement best describes the author's stand on Aristotle's idea of "good spirit"?

- o Agrees that long-term fulfilment is preferable to momentary happiness
- o Disagrees that the pursuit of happiness is the primary purpose of human life
- o Disagrees that fighting for a purpose leads to long-term fulfilment
- o Agrees that happiness and well-being are the ultimate goals of human life

Constructed Response Items: Main Idea

What was the main idea of this passage? That is, what was the author trying to communicate? Justify your response.

_____

_____

_____

_____

Constructed Response Items: Vocabulary

Directions. Provide a definition for each of the words displayed. If you are not certain of the definition, share what you think the meaning of the word may be.

Q1. warped

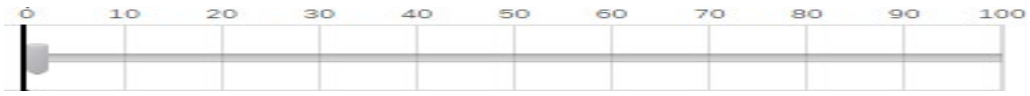
_____

Q2. masochistic

_____

Calibration of Performance

Direction. On a scale of 0 to 100, rate your performance on the comprehension questions you just answered (0 –very poor; 100 – excellent):



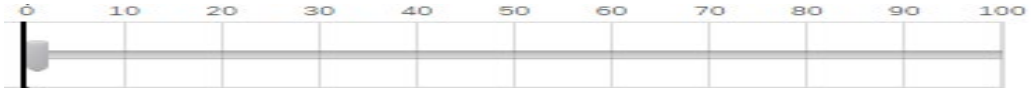
APPENDIX D: Post-Task Completion

Task Experience Measure

Direction. *On a scale of 0 to 100, rate how enjoyable and how difficult you found this experience:*

ENJOYMENT of the experience (0 – not enjoyable at all; 100 – highly enjoyable)

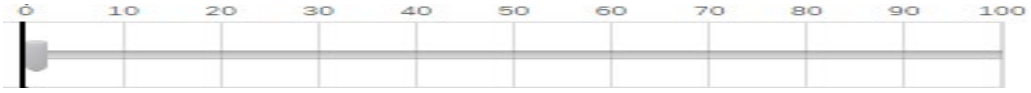
- *In Print*



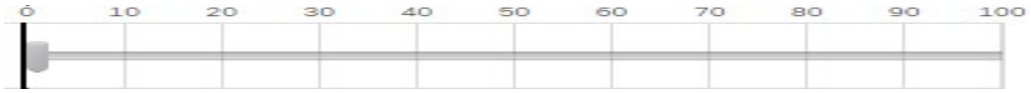
- *In Audio*

DIFFICULTY of the experience (0 – very easy; 100 – highly difficult)

- *In Print*



- *In Audio*



APPENDIX E: Cued Retrospective Interview Protocol and Text Segments

Instructions for Interview Facilitator:

1. Pre-select video segments to play before the interview
2. The video selections should include:
    a. Experimental Text excerpts (audio and print selections are included below)
3. Record on zoom cloud.
    a. To share sound --> click More --> select Share sound
    b. Play the audio and print segments in their entirety

4. Make the participant feel comfortable:
    "*We are here to better understand the differences in audio and print reading. It's not about your performance/end result but learning about the comprehension processes. So be open and candid – you are the expert!"*

5. The interview facilitator will read out the following prompt to the participant:
    "I will be showing you short videos of you completing the task. I will pause after playing a segment and ask you to explain what you were doing or thinking at that point. Verbalize everything that comes to mind."

6. If the participant stops talking for 5s, prompt them by saying:
    "*Is there anything else you would like to add about what you were doing?"*
    "Do you want to add anything more about what you were doing, thinking, or feeling?
    "Can you tell us more about what was going on in your head?"

7. Conduct the interview as a cued think aloud. Encourage participants to verbalize but do NOT put words in their mouth. Be comfortable with silence.

Experimental Texts

Segment 1: Inequality

5:15–6:36 The Spirit Level theory has been called "the left's new theory of everything," and it is as problematic as any other theory that leaps from a tangle of correlations to a single-cause explanation. For one thing, it's not obvious that people are whipped into competitive anxiety by the existence of J. K. Rowling and Sergey Brin as opposed to their own, local rivals for professional, romantic, and social success. Worse, economically egalitarian countries like Sweden and France differ from lopsided countries like Brazil and South Africa in many ways other than their income distribution. The egalitarian countries are, among other things, richer, better educated, better governed, and more culturally homogeneous, so a raw correlation between inequality and happiness (or any other social good) may show only that there are many reasons why it's better to live

in Denmark than in Uganda. Wilkinson and Pickett's sample was restricted to developed countries, but even within that sample the correlations are evanescent, coming and going with choices about which countries to include. Wealthy but unequal countries, such as Singapore and Hong Kong, are often socially healthier than poorer but more equal countries, such as those of ex-Communist Eastern Europe.

Segment 2: Happiness 1

0:42–2:05 When I read things like, "The foundations of capitalism are shattering," I'm like, maybe we need some time where we're walking around with a donkey with pots clanging on the sides. . . . 'Cause now we live in an amazing world, and it's wasted on the crappiest generation of spoiled idiots. . . . Flying is the worst one, because people come back from flights, and they tell you their story. . . . They're like, "It was the worst day of my life. . . . We get on the plane and they made us sit there on the runway for forty minutes." . . . Oh really, then what happened next? Did you fly through the air, incredibly, like a bird? Did you soar into the clouds, impossibly? Did you partake in the miracle of human flight, and then land softly on giant tires that you couldn't even conceive how they fuckin' put air in them? . . . You're sitting in a chair in the sky. You're like a Greek myth right now! . . . People say there's delays? . . . Air travel's too slow? New York to California in five hours. That used to take thirty years! And a bunch of you would die on the way there, and you'd get shot in the neck with an arrow, and the other passengers would just bury you and put a stick there with your hat on it and keep walking. . . . The Wright Brothers would kick us all in the [crotch] if they knew.

Segment 3: Happiness 2

4:06–5:35 Emotions and evaluations are, of course, related, though imperfectly: an abundance of happiness makes for a better life, but an absence of worry and sadness does not. And this brings us to the final dimension of a good life, meaning and purpose. This is the quality that, together with happiness, goes into Aristotle's ideal of eudaemonia or "good spirit." Happiness isn't everything. We can make choices that leave us unhappy in the short term but fulfilled over the course of a life, such as raising a child, writing a book, or fighting for a worthy cause.

Though no mortal can stipulate what really makes a life meaningful, the psychologist Roy Baumeister and his colleagues probed for what makes people feel their lives are meaningful. The respondents separately rated how happy and how meaningful their lives were, and they answered a long list of questions about their thoughts, activities, and circumstances. The results suggest that many of the things that make people happy also make their lives meaningful, such as being connected to others, feeling productive, and not being alone or bored. But other things can make lives happier while leaving them no more meaningful or even less so. People who lead happy but not necessarily meaningful lives have all their needs satisfied: they are healthy, have enough money, and feel good a lot of the time.

REFERENCES

Ackerman, R., & Goldsmith, M. (2011). Metacognitive regulation of text learning: On screen versus on paper. *Journal of Experimental Psychology: Applied*, *17*(1), 18–32. https://doi.org/10.1037/a0022086

Adair, J. G. (1984). The Hawthorne effect: A reconsideration of the methodological artifact. *Journal of Applied Psychology, 69*(2), 334–345. https://doi.org/10.1037/0021-9010.69.2.334

Afflerbach, P. (1986). The influence of prior knowledge on expert readers' importance assignment processes. *National Reading Conference Yearbook, 35*, 30–40.

Afflerbach, P., Cho, B. Y., & Kim, J. Y. (2015). Conceptualizing and assessing higher-order thinking in reading. *Theory Into Practice, 54*(3), 203–212. https://doi.org/10.1080/00405841.2015.1044367

Afflerbach, P., & Johnston, P. (1984). On the use of verbal reports in reading research. *Journal of Reading Behavior, 16*(4), 307–322.

Afflerbach, P., Pearson, D., & Paris, S. (2008). Clarifying differences between reading skills and reading strategies. *The Reading Teacher, 61*, 364–373. https://doi.org/10.1598/RT.61.5.1

Alexander, P.A. (1997) Mapping the multidimensional nature of domain learning: The interplay of cognitive, motivational, and strategic forces. In M.L. Maehr, & P.R. Pintrich (Eds.), *Advances in motivation and achievement* (Vol. 10, pp. 213–250). JAI Press.

Alexander, P.A. (2004). A model of domain learning: Reinterpreting expertise as a multidimensional, multistage process. In D.Y. Dai,

& R.J. Sternberg (Eds.), *Motivation, emotion, and cognition: Integrative perspectives on intellectual functioning and development* (pp. 273–298). Erlbaum.

Alexander, P.A. (2005). The path to competence: A lifespan developmental perspective on reading. *Journal of Literacy Research, 37*(4), 413–436. https://doi.org/10.1207/s15548430jlr3704_1

Alexander, P. A. (2019). The art (and science) of seduction: Why, when, and for whom seductive details matter. *Applied Cognitive Psychology, 33*(1), 142–148. https://doi.org/10.1002/acp.3510

Alexander, P. A., Jablansky, S., Singer, L. M., & Dumas, D. (2016). Relational reasoning: What we know and why it matters. *Policy Insights from The Behavioral and Brain Sciences*, *3*(1), 36–44. https://doi.org/10.1177/2372732215622029

Alexander, P. A., & Jetton, T. L. (2000). Learning from text: A multidimensional and developmental perspective. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 285–310). Lawrence Erlbaum Associates.

Alexander, P. A., Kulikowich, J. M., & Schulze, S. K. (1994). How subject-matter knowledge affects recall and interest. *American Educational Research Journal, 31*(2), 313–337. https://doi.org/10.3102/00028312031002313

Alexander, P. A., Murphy, P. K., & Kulikowich, J. M. (1998). What responses to domain-specific analogy problems reveal about emerging competence: A new perspective on an old acquaintance. *Journal of Educational Psychology, 90*(3),

397–406. https://doi.org/10.1037/0022-0663.90.3.397

Alexander, P. A., & the Disciplined Reading and Learning Research Laboratory. (2012). Reading into the future: Competence for the 21st century. *Educational Psychologist*, *47*(4), 259–280. https://doi.org/10.1080/00461520.2012.722511

Alexander, P. A., & the Disciplined Reading and Learning Research Laboratory (2020). Relational Reasoning: The bedrock of integration within and across multiple representations, documents, and perspectives. In P. van Meter, A. List, D. Lombardi, & P. Kendeou (Eds.)*, Handbook of learning from multiple representations and perspectives* (pp. 401–424). Routledge.

Anderson, R. C., Wilson, P. T., & Fielding, L. G. (1988). Growth in reading and how children spend their time outside of school. *Reading Research Quarterly*, *23*(3), 285–303. https://doi.org/10.1598/rrq.23.3.2

Andrade, J. (2010). What does doodling do?.*Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 24*(1), 100–106. https://doi.org/10.1002/acp.1561

Arvey, R. D., Strickland, W., Drauden, G., & Martin, C. (1990). Motivational components of test taking. *Personnel Psychology, 43*(4), 695–716. https://doi.org/10.1111/j.1744-6570.1990.tb00679.x

AudioFile (2023). Spotlight on popular narrators: Arthur Morey. Retrieved from https://www.audiofilemagazine.com/narrators/arthur-morey/

Audio Publishers Association. (2021, June 1). Audiobook sales rose 12% in 2020. [Press release]. Retrieved from https://www.audiopub.org/press/press

Baker, L., & Brown, N. (1984) Metacognitive skills and reading. In P.D. Pearson (Ed.), *Handbook of reading research* (pp. 353–394). Longman.

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis* (2nd ed.). Cambridge University Press. http://doi.org/10.1017/cbo9780511527685

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* Freeman.

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (Vol. 5, pp.307–337). Information Age Publishing.

Baron, N. (2021). *How we read now.* Oxford University Press.

Bartlett F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge University Press.

Bauer, E., Compton-Lilly, C., Li, G., & Razfar, A. (2021). Centering learners: Literate identities and literacy education. *Journal of Literacy Research*, *53*(4), 435–436. https://doi.org/10.1177/1086296X211052453

Bauer, D. J., & Curran, P. J. (2004). The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. *Psychological Methods, 9*(1), 3–29. https://doi.org/10.1037/1082-989X.9.1.3

Benjamin Jr, L. T., Cavell, T. A., & Shallenberger III, W. R. (1984). Staying with initial answers on objective tests: Is it a myth?. *Teaching of Psychology, 11*(3), 133–141.

Berger, N. S., & Perfetti, C. A. (1977). Reading skill and memory for spoken and written discourse. *Journal of Reading Behavior*, *9*(1), 7–16. https://doi.org/10.1080/10862967709547200

Bernstein, M. R. (1955). Relationship between interest and reading comprehension. *The Journal of Educational* Research, *49*(4), 283–288. https://doi.org/10.1080/00220671.1955.10882283

Bishop, D.V., & Adams, C. (1992). Comprehension problems in children with specific language impairment: Literal and inferential meaning. *Journal of Speech and Hearing Research, 35*(1), 119–129. https://doi.org/10.1044/jshr.3501.119

Bol, L., & Hacker, D. J. (2012). Calibration research: Where do we go from here? *Frontiers in Psychology, 3*, 1–6. https://doi.org/10.3389/fpsyg.2012.00229

Bolck, A., Croon, M., & Hagenaars, J. (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political Analysis, 12*, 3–27. https://doi.org/10.1093/pan/mph001

Bookheimer, S. (2002). Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. *Annual Review of Neuroscience*, *25*, 151–188. https://doi.org/10.1146/annurev.neuro.25.112701.142946

Boyle, E. A., Rosenberg, M. S., Connelly, V. J., Washburn, S. G., Brinckerhoff, L. C., & Banerjee, M. (2003). Effects of audio texts on the acquisition of secondary–level content by students with mild disabilities. *Learning Disability Quarterly*, *26*(3), 203–214 . https://doi.org/10.2307/1593652

Bransford, J., & Johnson, M. K. (1972). Contextual prerequisites for understanding some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11*, 717–726.

Braze, D., Mencl, W. E., Tabor, W., Pugh, K. R., Todd Constable, R., Fulbright, R. K., . . . Shankweiler, D. P. (2011). Unification of sentence processing via ear and eye: An fMRI study. *Cortex*, *47*(4), 416–416. https://doi.org/10.1016/j.cortex.2009.11.005

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading–while–listening, and listening to stories. *Reading in a Foreign Language, 20*(2), 136–163.

Buchanan, J., Summerville, A., Lehmann, J., & Reb, J. (2016). The Regret Elements Scale: Distinguishing the affective and cognitive components of regret. *Judgment and Decision Making, 11*(3), 275–286.

Buchweitz, A., Mason, R. A., Tomitch, L. M. B., Just, M. A., & Hall, A. B. B. (2009). Brain activation for reading and listening comprehension: An fMRI study of modality effects and individual differences in language comprehension. *Psychology & Neuroscience*, *2*(2), 111–123. https://doi.org/10.3922/j.psns.2009.2.003

Cain, K., Oakhill, J. V., Barnes, M. A., Bryant, P. E. (2001). Comprehension skill, inference making ability, and the relation to knowledge. Memory & Cognition, 29, 850–859. https://doi.org/10.3758/BF03196414

Caracelli, V. J., & Greene, J. C. (1993). Data analysis strategies for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis, 15*(2), 195–207.

Carriere, J. S., Seli, P., & Smilek, D. (2013). Wandering in both mind and body: Individual differences in mind wandering and inattention predict fidgeting. *Canadian Journal of Experimental Psychology, 67*, 19–31. https://doi.org/10.1037/a0031438 https://doi.org/10.3102/01623737015002195

Chall, J. S. (1983). *Stages of reading development*. McGraw-Hill.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*(2), 121–152. https://doi.org/10.1207/s15516709cog0502_2

Clark, V. L. P. (2019). Meaningful integration within mixed methods studies: Identifying why, what, when, and how. *Contemporary Educational Psychology, 57*, 106 – 111. https://doi.org/10.1016/j.cedpsych.2019.01.007

Clinton, V. (2019). Reading from paper compared to screens: A systematic review and meta–analysis. *Journal of Research in Reading*, *42*(2), 288–325. https://doi.org/10.1111/1467-9817.12269

Clinton, V., Taylor, T., Bajpayee, S., Davison, M. L., Carlson, S. E., & Seipel, B. (2020). Inferential comprehension differences between narrative and expository texts: a systematic review and meta–analysis. *Reading and Writing*, *33*(9), 2223–2248. https://doi.org/10.1007/s11145–020–10044–2/tables/4

Clinton-Lisell, V. (2021). Listening ears or reading eyes: A meta-analysis of reading and listening comprehension comparisons. Review of Educational Research. https://doi.org/10.3102/00346543211060871

Coiro, J., & Dobler, E. (2007). Exploring the online reading comprehension strategies used by sixth-grade skilled readers to search for and locate information on the

Internet. *Reading Research Quarterly, 42*(2), 214 –257.

Constable, R. T., Pugh, K. R., Berroya, E., Mencl, W. E., Westerveld, M., Ni, W., & Shankweiler, D. (2004). Sentence complexity and input modality effects in sentence comprehension: An fMRI study. *NeuroImage*, *22*(1), 11–21. https://doi.org/10.1016/j.neuroimage.2004.01.001

Couchman, J. J., Miller, N. E., Zmuda, S. J., Feather, K., & Schwartzmeyer, T. (2016). The instinct fallacy: The metacognition of answering and revising during college exams. *Metacognition and Learning, 11*(2), 171–185. https://doi.org/10.1007/s11409-015-9140-8

Creswell, J. W. (2009). Mapping the field of mixed methods research [Editorial]. *Journal of Mixed Methods Research, 3*(2), 95–108. https://doi.org/10.1177/1558689808330883

Creswell, J. W., Plano Clark, V. L., Gutmann, M. L., & Hanson, W. E. (2003). Advanced mixed methods research designs. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 209–240). Sage.

Cunningham, A. E., & Stanovich, K. E. (1991). Tracking the unique effects of print exposure in children: associations with vocabulary, general knowledge, and spelling. *Journal of Educational Psychology*, *83*(2), 264–274. https://doi.org/10.1037/0022–0663.83.2.264

Danckert, J., Hammerschmidt, T., Marty-Dugas, J., & Smilek, D. (2018). Boredom: Under-aroused and restless. *Consciousness and Cognition, 61*, 24–37. https://doi.org/10.1016/j.concog.2018.03.014

Daniel, D. B., & Woody, W. D. (2010). They hear, but do not listen: Retention for podcasted material in a classroom context. *Teaching of Psychology*, *37*(3), 199–203. https://doi.org/10.1080/00986283.2010.488542

Dehaene, S. (2014). Evolution of human cortical circuits for reading and arithmetic: The "neuronal recycling" hypothesis. In S. Dehaene, J. R. Duhamel, M. Hauser, & G. Rizzolatti (Eds.), *From monkey brain to human brain* (pp. 133–157). MIT Press.

Dehaene, S. (2020). *How we learn: Why brains learn better than any machine…for now.* Penguin Random House.

Delgado, P., Vargas, C., Ackerman, R., & Salmerón, L. (2018). Don't throw away your printed books: A meta–analysis on the effects of reading media on reading comprehension. *Educational Research Review*, *25*, 23–38. https://doi.org/10.1016/j.edurev.2018.09.003

DeMars, C. E. (2000). Test stakes and item format interactions. *Applied Measurement in Education, 13*(1), 55–77. https://doi.org/10.1207/s15324818ame1301_3

Deniz, F., Nunez–Elizalde, A. O., Huth, A. G., & Gallant, J. L. (2019). The representation of semantic information across human cerebral cortex during listening versus reading is invariant to stimulus modality. *Journal of Neuroscience*, *39*(39), 7722–7736. https://doi.org/10.1523/jneurosci.0675–19.2019

Diakidoy, I. A. N., Stylianou, P., Karefillidou, C., & Papageorgiou, P. (2005). The relationship between listening and reading comprehension of different types of text at increasing grade levels. *Reading Psychology, 26*(1), 55–80. https://doi.org/10.1080/02702710590910584

Dinsmore, D. L., & Alexander, P. A. (2012). A critical discussion of deep and surface processing: What it means, how it is measured, the role of context, and model specification. *Educational Psychology Review, 24*(4), 499–567. https://doi.org/10.1007/s10648-012-9198-7

Dinsmore, D. L., Loughlin, S. M., Parkinson, M. M., & Alexander, P. A. (2015). The effects of persuasive and expository text on metacognitive monitoring and control. *Learning and Individual Differences*, *38*, 54–60. https://doi.org/10.1016/j.lindif.2015.01.009

Dinsmore, D. L., & Parkinson, M. M. (2013). What are confidence judgments made of? Students' explanations for their confidence ratings and what that means for calibration. *Learning and Instruction*, *24*(1), 4–14. https://doi.org/10.1016/j.learninstruc.2012.06.001

Dochy, F., Segers, M., & Buehl, M. M. (1999). The Relation between assessment practices and outcomes of studies: The case of research on prior knowledge. *Review of Educational Research*, *69*(2), 145–145. https://doi.org/10.2307/1170673

Dudycha, A. L., & Carpenter, J. B. (1973). Effects of item format on item discrimination and difficulty. *Journal of Applied Psychology, 58*(1), 116–121. https://doi.org/10.1037/h0035197

Duker, S. (1965). Listening and reading. *The Elementary School Journal*, *65*(6), 321–329. https://doi.org/10.1086/460227

Dunkel, P. (1991). Listening in the native and second/foreign language: Toward an integration of research and practice. *TESOL Quarterly, 25*(3), 431–457.

Dunlosky, J., & Rawson, K. A. (2010). Why does rereading improve metacomprehension accuracy? Evaluating the levels-of-disruption hypothesis for the rereading effect. *Discourse Processes*, *40*(1), 37–55. https://doi.org/10.1207/S15326950DP4001_2

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53*(1), 109–132. https://doi.org/10.1146/annurev.psych.53.100901.135153

Edison Research & Triton Digital. (2021). *The infinite dial 2021*. http://www.edisonresearch.com/wp-content/uploads/2021/03/The-Infinite-Dial-2021.pdf

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading, 9*(2), 167–188. https://doi.org/10.1207/s1532799xssr0902_4

Ehri, L. C., & Sweet, J. (1991). Fingerpoint-reading of memorized text: What enables beginners to process the print? *Reading Research Quarterly, 26*(4), 442–462. https://doi.org/10.2307/747897

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal Of Advanced Nursing, 62*(1), 107–115. https://doi.org/10.1111/j.1365-2648.2007.04569.x

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*(3), 215–250. https://doi.org/10.1.1.697.3088&rep=rep1&type=pdf

Farley, J., Risko, E., & Kingstone, A. (2013). Everyday attention and lecture retention: The effects of time, fidgeting, and mind wandering. *Frontiers in Psychology, 4*, 619.

Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: A review of the role of

motor system in language comprehension. *The Quarterly Journal of Experimental Psychology, 61*(6), 825–850. https://doi.org/10.1080/17470210701623605

Fletcher, C. R., & Bloom, C. P. (1988). Causal reasoning in the comprehension of simple narrative texts. *Journal of Memory and Language, 27*(3), 235–244. https://doi.org/10.1016/0749-596X(88)90052-6

Fletcherand, C. R., & Bloom, P. (1988). Causal reasoning in the comprehension of simple narrative texts. *Journal of Memory and Language*, *21*, 235–244.

Fox, E., & Alexander, P. A. (2009). Text comprehension: A retrospective, perspective, and prospective. In S. E. Israel & G. G. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. 227–239). Routledge.

Freebody, P., & Luke, A. (1990). Literacies programs: Debates and demands in cultural context. *Prospect: An Australian Journal of TESOL*, *5*(3), 7–16.

Freeman, M. (2016). *Historicising transmedia storytelling: Early twentieth-century transmedia story worlds*. Routledge.

Friedland, A., Gilman, M., Johnson, M., & Demeke, A. (2017). Does reading-while-listening enhance students' reading fluency? Preliminary results from school experiments in rural Uganda. *Journal of Education and Practice, 8*(7), 82–95.

Fuchs, L.S., Fuchs, D., Hosp, M.K. & Jenkins, J.R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5*(3), 239–256. https://doi.org/10.1207/s1532799xssr0503_3

Garner, R. (1987). *Metacognition and reading comprehension*. Ablex Publishing.

Garner, R., & Alexander, P. A. (1989). Metacognition: Answered and unanswered

questions. *Educational Psychologist, 24*(2), 143–158. https://doi.org/10.1207/s15326985ep2402_2

Garner, R., Alexander, P. A., Gillingham, M. G., Kulikowich, J. M., & Brown, R. (1991). Interest and learning from text. *American Educational Research Journal, 28*(3), 643–659. https://doi.org/10.2307/1163152

Garner, R., Hare, V. C., Alexander, P., Haynes, J., & Winograd, P. (1984). Inducing use of a text lookback strategy among unsuccessful readers. *American Educational Research Journal, 21*(4), 789–798. https://doi.org/10.3102/00028312021004789

Garner, R., & Kraus, C. (1981). Good and poor comprehender differences in knowing and regulating reading behaviors. *Educational Research Quarterly, 6*(4), 5–12.

Geiser, C., Lehmann,W., & Eid, M. (2006). Separating "rotators" from "non-rotators" in the mental rotations test: A multigroup latent class analysis. *Multivariate Behavioral Research, 41*(3), 261–293. https://doi.org/10.1207/s15327906mbr4103_2

Gernsbacher, M. A., Varner, K. R., & Faust, M. E. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(3), 430–445. https://doi.org/10.1037/0278-7393.16.3.430

Glenberg, A. M., & Epstein, W. (1985). Calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(4), 702–718. https://doi.org/10.1037/0278–7393.11.1-4.702

Glenberg, A. M., Sanocki, T., Epstein, W., & Morris, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General*, *116*(2), 119–136.

https://doi.org/10.1037/0096-3445.116.2.119

Goh, C. C. (2000). A cognitive perspective on language learners' listening comprehension problems. *System, 28*(1), 55–75. http://doi.org/10.1016/s0346–251x(99)00060-3

Gopher, D., & Donchin, E. (1986). Workload: An examination of the concept. In K.Boff & L. Kaufman (Eds.), *Handbook of human perception and performance,* 1–49. John Wiley.

Gough, P.B. & Tunmer, W.E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*(1), 6–10. https://doi.org/10.1177/074193258600700104

Grace Kim, Y.–S., & Petscher, Y. (2020). Influences of individual, text, and assessment factors on text/discourse comprehension in oral language (listening comprehension). *Annals of Dyslexia, 71,* 218–237. https://doi.org/10.1007/s11881-020-00208-8

Grant, D. A. (1948). The Latin square principle in the design and analysis of psychological experiments. *Psychological Bulletin*, *45*(5), 427–442. https://doi.org/10.1037/h0053912

Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed methods evaluation designs. *Educational Evaluation and Policy Analysis, 11*(3), 255–274. https://doi.org/10.3102/01623737011003255

Guthrie, J. T., & Wigfield, A. (2000). Engagement and motivation in reading. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research, Vol. 3* (pp. 403–422). Lawrence Erlbaum Associates.

Guthrie, J. T., Wigfield, A., Metsala, J. L., & Cox, K. E. (2009). Motivational and

cognitive predictors of text comprehension and reading amount. *Scientific Studies of Reading,* 3(3), 231–256. https://doi.org/10.1207/S1532799XSSR0303_3

Gutierrez, A. P., & Schraw, G. (2014). Effects of strategy training and incentives on students' performance, confidence, and calibration. *The Journal of Experimental Education,* 83(3), 384–404. https://doi.org/10.1080/00220973.2014.907230

Haas, C. (1989). Does the medium make a difference? Two studies of writing with pen and paper and with computers. *Human-Computer Interaction*, *4*(2), 149 –169.

Haenggi, D., & Perfetti, C. A. (1992). Individual differences in reprocessing of text. *Journal of Educational Psychology*, *84*(2), 182–192. https://doi.org/10.1037/0022-0663.84.2.182

Haenggi, D., & Perfetti, C. A. (1994). Processing components of college-level reading comprehension. *Discourse Processes*, *17*(1), 83–104. https://doi.org/10.1080/01638539409544860

Harmes, J. C., & Wise, S. L. (2016). Assessing engagement during the online assessment of real-world skills. In Y. Rosen, S. Ferrara, & M. Mosharraf (Eds.) *Handbook of research on technology tools for real-world skill development* (pp. 805–824). IGI Global.

Have, I., & Stougaard Pedersen, B. (2016). *Digital audiobooks: New media, users, and experiences*. Routledge.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. T*he Physics Teacher, 30*(3), 141–158.

Hirsch, E. D., Jr. (2003). Reading comprehension requires knowledge—Of words and the world: Scientific insights into the fourth-grade slump and the nation's stagnant

comprehension scores. *American Educator, 1*, 10–45. https://doi.org/10.1.1.141.6106

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal, 2*(2), 127–160. https://doi.org/10.1007/BF00401799

Horkay, N., Bennett, R. E., Allen, N., Kaplan, B., & Yan, F. (2006). Does it matter if I take my writing test on computer? An empirical study of mode effects in NAEP. *Journal of Technology, Learning, and Assessment, 5*(2). http://www.jtla.org

Huang, S., Capps, M., Blacklock, J., & Garza, M. (2014). Reading habits of college students in the United States. *Reading Psychology*, *35*(5), 437–467. https://doi.org/10.1080/02702711.2012.739593

Jenkins, J. J. (1979). Four points to remember: A tetrahedral model of memory experiments. In L. S. Cermak & F. L. Craik (Eds.), *Levels of processing in human memory* (pp. 429–446). Lawrence Erlbaum Associates.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329–354. https://doi.org/10.1037/0033-295X.87.4.329

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

Keller, T. A., Carpenter, P. A., & Just, M. A. (2001). The neural bases of sentence comprehension: A fMRI examination of syntactic and lexical processing. *Cerebral Cortex*, *11*(3), 223–237. https://doi.org/10.1093/cercor/11.3.223

Kendeou, P., Bohn–Gettler, C., White, M. J., & Van Den Broek, P. (2008). Children's inference generation across different media. *Journal of Research in Reading*,

31(3), 259–272. https://doi.org/10.1111/j.1467-9817.2008.00370.x

Kendeou, P., Butterfuss, R., Van Boekel, M., & O'Brien, E. J. (2017). Integrating relational reasoning and knowledge revision during reading. *Educational Psychology Review, 29*(1), 27–39. https://doi.org/10.1007/s10648-016-9381-3

Kendeou, P., Lynch, J. S., van den Broek, P., Espin, C. A., White, M. J., & Kremer, K. E. (2005). Developing successful readers: Building early comprehension skills through television viewing and listening. *Early Childhood Education Journal*, *33*(2), 91–98. https://doi.org/10.1007/s10643–005–0030–6

Kendeou, P., & O'Brien, E. J. (2018). Reading comprehension theories: A view from the top down. In M. F. Schober, D. N. Rapp, & M. A. Britt (Eds.), *The Routledge handbook of discourse processes* (pp. 7–21). Routledge/Taylor & Francis Group.

Kendeou, P., & van den Broek, P. (2007). The effects of prior knowledge and text structure on comprehension processes during reading of scientific texts. *Memory & Cognition*, *35*(7), 1567–1577. https://doi.org/10.3758/bf03193491

Kendeou, P., Van Den Broek, P., Helder, A., & Karlsson, J. (2014). A cognitive view of reading comprehension: Implications for reading difficulties. *Learning Disabilities Research & Practice*, *29*(1), 10–16. https://doi.org/10.1111/ldrp.12025

Kendeou, P., van den Broek, P., White, M. J., Lynch, J. S. (2009). Predicting reading comprehension in early elementary school: The independent contributions of oral language and decoding skills. *Journal of Educational Psychology, 101*, 765 –778.

Kim, Y. S. G., & Petscher, Y. (2021). Influences of individual, text, and assessment factors on text/discourse comprehension in oral language (listening

comprehension). *Annals of Dyslexia, 71*(2), 218–237. https://doi.org/10.1007/s11881-020-00208-8

Kintsch, W. (1979, April 8–12). *On comprehension* [Paper presentation]. American Educational Research Association Annual Meeting, San Francisco, CA, United States.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*, 163–182.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.

Kintsch, W. (2018). Revisiting the construction-integration model of text comprehension and its implications for instruction. In D.E. Alvermann, N.J. Unrau, M. Soliers, & R.B. Ruddell (Eds.), *Theoretical models and processes of literacy* (7th ed., pp. 178–203). Routledge. https://doi.org/10.4324/9781315110592-12

Kintsch, W., & Kozminsky, E. (1977). Summarizing stories after reading and listening. *Journal of Educational Psychology, 69*(5), 491–499. https://doi.org/10.1037/0022–0663.69.5.491

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*(5), 363–394. https://doi.org/10.1037/0033-295X.85.5.363

Kintsch, W., & Young, S. R. (1984). Selective recall of decision-relevant information from texts. *Memory & Cognition, 12*(2), 112–117. https://doi.org/10.3758/BF03198424

Kline, R. B. (2012) Assumptions in structural equation modeling. In R. H. Hoyle (Ed.),

Handbook of structural equation modeling (pp.111–125). Guilford Press. https://doi.org/10.1080/10705511.2013.769397

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *Modern Language Journal, 73*(4), 440–464. https://doi.org/10.2307/326879

Kruger, J., Wirtz, D., & Miller, D. T. (2005). Counterfactual thinking and the first instinct fallacy. *Journal of Personality and Social Psychology, 88*(5), 725–735. https://doi.org/10.1037/0022-3514.88.5.725

Kucan, L., & Beck, I. L. (1997). Thinking aloud and reading comprehension research: Inquiry, instruction, and social interaction. *Review of Educational Research*, *67*(3), 271–299. https://doi.org/10.3102/00346543067003271

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*(2), 293–323. https://doi.org/10.1016/0010-0285(74)90015-2

Labuhn, A. S., Zimmerman, B. J., & Hasselhorn, M. (2010). Enhancing students' self-regulation and mathematics performance: The influence of feedback and self-evaluative standards. *Metacognition and Learning*, *5*(2), 173–194. https://doi.org/10.1007/s11409-010-9056-2

Langner, R., & Eickhoff, S. B. (2013). Sustaining attention to simple tasks: A meta-analytic review of the neural mechanisms of vigilant attention. *Psychological Bulletin, 139*(4), 870–900. https://doi.org/10.1037/a0030694

Leahy, W., & Sweller, J. (2011). Cognitive load theory, modality of presentation and the transient information effect. *Applied Cognitive Psychology*, *25*(6), 943–951.

https://doi.org/10.1002/acp.1787

Lee, Y. H., & Jia, Y. (2014). Using response time to investigate students' test-taking behaviors in a NAEP computer-based study. *Large-scale Assessments in Education, 2*(1), 1–24. https://doi.org/10.1186/s40536-014-0008-1

Lehman, H. C. (1928). Does it pay to change initial decisions in a true-false test?. *School & Society*, *28,* 456–458.

Lewandowski, L., Gathje, R. A., Lovett, B. J., & Gordon, M. (2013). Test-taking skills in college students with and without ADHD. *Journal of Psychoeducational Assessment, 31*(1), 41–52.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: the state of the art to 1980. In D. Kahneman, P. Slavic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306 – 334). Cambridge University Press.

List, A., & Alexander, P. A. (2017). Analyzing and integrating models of multiple text comprehension. *Educational Psychologist, 52*(3), 143–147. https://doi.org/10.1080/00461520.2017.1328309

List, A., & Ballenger, E. E. (2019). Comprehension across mediums: The case of text and video. *Journal of Computing in Higher Education*. https://doi.org/10.1007/s12528-018-09204-9

Liu, Z. Reading behavior in the digital environment: Changes in reading behavior over the past ten years. *Journal of Documentation*, *61*(6), 700–712. https://doi.org/10.1108/00220410510632040

Lowe, M. L., & Crawford, C. C. (1929). First impression versus second thought in true-

false tests. *Journal of Educational Psychology*, 20(3), 192–195.

Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology, 9*(1), 111–151. https://doi.org/10.1016/0010-0285(77)90006-8

Mangen, A., & Van der Weel, A. (2016). The evolution of reading in the age of digitisation: An integrative framework for reading research. *Literacy*, *50*(3), 116 – 124. https://doi.org/10.1111/lit.12086

Mayer, R. (2005). Cognitive theory of multimedia learning. In Mayer, R. (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31–48). Cambridge University Press.

McCambridge, J., Witton, J., & Elbourne, D. R. (2014). Systematic review of the Hawthorne effect: N concepts are needed to study research participation effects. *Journal of Clinical Epidemiology, 67*(3), 267–277. https://doi.org/10.1016/j.jclinepi.2013.08.015

McCrudden, M. T., & Marchand, G. (2020). Multilevel mixed methods research and educational psychology. *Educational Psychologist*, *55*(4), 197–207. https://doi.org/10.1080/00461520.2020.1793156

McDaniel, M. A., & Donnelly, C. M. (1996). Learning with analogy and elaborative interrogation. *Journal of Educational Psychology, 88*(3), 508–519. https://doi.org/10.1037/0022-0663.88.3.508

McGeown, S. P., Duncan, L.G., Griths, Y. M., & Stothard, S.E. (2015) Exploring the relationship between adolescent's reading skills, reading motivation and reading habits. *Reading and Writing: An Interdisciplinary Journal, 28*(4), 545 –569.

https://doi.org/10.1007/s11145-014-9537-9

McLuhan, M. (1964). *Understanding media: The extensions of man*. McGraw-Hill College.

McMaster, K. L., & Kendeou, P. (2023). Refocusing reading comprehension: Aligning theory with assessment and intervention. *Learning and Individual Differences,* 102256. https://doi.org/10.1016/j.lindif.2023.102256

McMorris, R. F., DeMers, L. P., & Schwarz, S. P. (1987). Attitudes, behaviors, and reasons for changing responses following answer-changing instruction. *Journal of Educational Measurement, 24*(2), 131–143. https://doi.org/10.1111/j.1745-3984.1987.tb00269.x

McNamara, D. S. (2001). Reading both high-coherence and low-coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology*, *55*(1), 51–62. https://doi.org/10.1037/H0087352

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, *14*(1), 1–43. https://doi.org/10.1207/s1532690xci1401_1

McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes*, *22*(3), 247–288. https://doi.org/10.1080/01638539609544975

McNamara, D. S., & Magliano, J. (2009). Toward a comprehensive model of comprehension. *Psychology of learning and motivation, 51*, 297–384. https://doi.org/10.1016/S0079-7421(09)51009-2

McNamara, D. S., & Shapiro, A. M. (2005). Multimedia and hypermedia solutions for

promoting metacognitive engagement, coherence, and learning. *Journal of Educational Computing Research, 33*(1): 1–29. https://doi.org/10.2190/7N6R-PCJL-UMHK-RYPJ

Meyer, B.J.F. (1975). Identification of the structure of prose and its implications for the study of reading and memory. *Journal of Literacy Research, 7*(1), 7–47. https://doi.org/10.1080/10862967509547120

Mol, S. E., & Bus, A. G. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin*, *137*(2), 267–296. https://doi.org/10.1037/A0021890

Mokhtari, K., Reichard, C. A., & Gardner, A. (2009). The impact of internet and television use on the reading habits and practices of college students. *Journal of Adolescent & Adult Literacy*, *52*(7), 609–619. https://doi.org/10.1598/JAAL.52.7.6

Morris, J. W., & Patterson, E. (2015). Podcasting and its apps: Software, sound, and the interfaces of digital audio. *Journal of Radio & Audio Media, 22*(2), 220–230. https://doi.org/10.1080/19376529.2015.1083374

Murphy, P. K., Long, J. F., Holleran, T. A., & Esterly, E. (2003). Persuasion online or on paper: A new take on an old issue. *Learning and Instruction, 13*(5), 511–532. https://doi.org/10.1016/S0959-4752(02)00041-5

Nakashima, K., Stephens, M., & Kamata, S. (2018). The interplay of silent reading, reading-while-listening and listening-only. *The Reading Matrix: An International Online Journal, 18*(1), 104–123.

National Assessment Governing Board. (2015). *Reading framework for the 2015 National Assessment of Educational Progress*. Washington, DC: Author.

National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). *Common core state standards (English language arts)*. Washington D.C.: Author. http://corestandards.org/

Nicholls, J. G. (1979). Development of perception of own attainment and causal attributions for success and failure in reading. *Journal of Educational Psychology, 71*(1), 94–99. https://doi.org/10.1037/0022-0663.71.1.94

Nietfeld, J. L., Cao, L., & Osborne, J. W. (2006). The effect of distributed monitoring exercises and feedback on performance, monitoring accuracy, and self-efficacy. *Metacognition and Learning, 1*(2), 159–179. https://doi.org/10.1007/S10409-006-9595-6

Nigro, G., & Neisser, U. (1983). Point of view in personal memories. *Cognitive Psychology, 15*(4), 467–482. https://doi.org/10.1016/0010-0285(83)90016-6

Noland, T. (2020, October 26). Helping students with dyslexia experience learning success. Learning Ally. https://learningally.org/Solutions-for-School/Educator-Blog/helping-students-with-dyslexia-experience-learning-success

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14*(4), 535–569. https://doi.org/10.1080/10705510701575396

Nylund-Gibson, K.L., Grimm, R.P., & Masyn K.E. (2019) Prediction from latent classes: A demonstration of different approaches to include distal outcomes in mixture

models. *Structural Equation Modeling: A Multidisciplinary Journal, 26*(6), 967–985, https://doi.org/10.1080/10705511.2019.1590146

Olshavsky, J. E. (1976, 1977). Reading as problem solving: An investigation of strategies. *Reading Research Quarterly*, *12*(4), 654–654. https://doi.org/10.2307/747446

Olson, R. K., Kliegl, R., & Davidson, B. J. (1983). Dyslexic and normal readers' eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, *9*(5), 816–825. https://doi.org/10.1037/0096-1523.9.5.816

Ozuru, Y., Dempsey, K., & McNamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction*, *19*(3), 228–242. https://doi.org/10.1016/j.learninstruc.2008.04.003

Parkinson, M. M., & Dinsmore, D. L. (2010). Calibrating calibration: towards conceptual clarity and agreement in calculation.

Paris, S. G., & Oka, E. R. (1986). Children's reading strategies, metacognition, and motivation. *Developmental Review, 6*(1), 25–56. https://doi.org/10.1016/0273-2297(86)90002-X

Pastor, D.A., & Erbacher, M.K. (2018) Cluster analysis. In G.R. Hancock, L.M. Stapleton, & R.O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences.* Routledge.

Pastor, D. A., Barron, K. E., Miller, B. J., & Davis, S. L. (2007). A latent profile analysis of college students' achievement goal orientation. *Contemporary Educational Psychology, 32*(1), 8–47. https://doi.org/10.1016/j.cedpsych.2006.10.003

Penno, J. F., Wilkinson, I. A. G., & Moore, D. W. (2002). Vocabulary acquisition from teacher explanation and repeated listening to stories: Do they overcome the Matthew effect? *Journal of Educational Psychology, 94*(1), 23-33. https://doi.org/10.1037/0022-0663.94.1.23

Perfetti, C. (1985). *Reading ability*. Oxford University Press.

Pew Research Center. (2019, September 25). One–in–five Americans now listen to audiobooks. Retrieved from https://www.pewresearch.org

Pew Research Center (2021, June 29). Audio and podcasting fact sheet. Retrieved from https://www.pewresearch.org

Pew Research Center (2022, February 15). Nearly a quarter of Americans get news from podcasts. Retrieved from https://www.pewresearch.org

Pieschl, S. (2008). Metacognitive calibration: An extended conceptualization and potential applications. *Metacognition and Learning*, *4*(1), 3–31. https://doi.org/10.1007/S11409-008-9030-4

Piolat, A., Roussey, J. Y., & Thunin, O. (1997). Effects of screen presentation on text reading and revising. *International Journal of Human-Computer Studies, 47*(4), 565–589.

Prinsell, C. P., Ramsey, P. H., & Ramsey, P. P. (1994). Score gains, attitudes, and behavior changes due to answer-changing instruction. *Journal of Educational Measurement, 31*(4), 327–337.

Proudfoot, K. (2022). Inductive/Deductive hybrid thematic analysis in mixed methods research. *Journal of Mixed Methods Research.* https://doi.org/10.1177/15586898221126816

Snow, C. (2002). Reading for understanding: Toward an R&D program in reading comprehension. RAND Corporation.

Rapp, D. N., Broek, P. V. D., McMaster, K. L., Kendeou, P., & Espin, C. A. (2007). Higher-order comprehension processes in struggling readers: A perspective for research and intervention. *Scientific Studies of Reading, 11*(4), 289–312. https://doi.org/10.1080/10888430701530417

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*(3), 191–201. https://doi.org/10.3758/BF03197692

Rios, J. (2021). Improving test-taking effort in low-stakes group-based educational testing: A meta-analysis of interventions. *Applied Measurement in Education, 34*(2), 85–106. https://doi.org/10.1080/08957347.2021.1890741

Robertson, I. H., & Garavan, H. (2004). Vigilant attention. In M. S. Gazzaniga (Ed.), The cognitive neurosciences (pp. 631–640). Boston Review.

Roland, D., Thoma, B., Tagg, A., Woods, J., Chan, T. M., & Riddell, J. (2021). What are the real-world podcast-listening habits of medical professionals? *Cureus, 13*(7):e16240. https://doi.org/10.7759/cureus.16240

Ronconi, A., Veronesi, V., Mason, L., Manzione, L., Florit, E., Anmarkrud, Ø., & Bråten, I. (2022). Effects of reading medium on the processing, comprehension, and calibration of adolescent readers. *Computers & Education*, 185, 104520. https://doi.org/10.1016/j.compedu.2022.104520

Rubery, M. (Ed.). (2011). *Audiobooks, literature, and sound studies*. Routledge.

Rubery, M. (2016). *The untold story of the talking book*. Harvard University Press.

Rumelhart, D. E. (1994). Toward an interactive model of reading. In. R.B. Ruddell, M.R. Ruddell, & H. Singer (Eds.), *Theoretical models and processes of reading* (pp. 864–894). International Reading Association.

Salmerón, L., Naumann, J., García, V., & Fajardo, I. (2017). Scanning and deep processing of

information in hypertext: An eye-tracking and cued retrospective think-aloud study. *Journal of Computer Assisted Learning, 33*, 222–233. https://doi.org/0.1111/jcal.12152

Sanchez, C. A., & Wiley, J. (2009). To scroll or not to scroll: Scrolling, working memory capacity, and comprehending complex texts. *Human Factors, 51*(5), 730–738. https://doi.org/10.1177/0018720809352788

Schiefele, U. (1992). Topic interest and levels of text comprehension. In K. A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 151–182). Lawrence Erlbaum Associates.

Schneider, S. L., & Laurion, S. K. (1993). Do we know what we've learned from listening to the news? *Memory & Cognition*, *21*(2), 198–209.

School-to-School International. (2017). Technology-based innovations to improve early grade reading outcomes in developing countries. All Children Reading: A Grand Challenge for Development (ACR GCD). https://allchildrenreading.org/wp-content/uploads/2019/07/Summative-EdTech-for_Literacy-Web-2.pdf

Schraw, G. (2009). Measuring metacognitive judgments. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 415 –429). Routledge.

Schraw, G., Potenza, M. T., & Nebelsick-Gullet, L. (1993). Constraints on the calibration of performance. *Contemporary Educational Psychology, 18*(4), 455–463. https://doi.org/10.1006/ceps.1993.1034

Schunk, D. H., & DiBenedetto, M. K. (2020). Motivation and social cognitive theory. *Contemporary Educational Psychology*, *60*, 101832–101832. https://doi.org/10.1016/j.cedpsych.2019.101832

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Scott, D. B. (2008). Assessing text processing: A comparison of four methods. *Journal of Literacy Research*, *40*, 290–316. https://doi.org/10.1080/10862960802502162

Seda, I., & Pearson, P. D. (2010). Interviews to assess learners' outcomes. *Reading Research and Instruction*, *31*(1), 22–32. https://doi.org/10.1080/19388079109558068

Sénéchal, M., Pagan, S., Lever, R., & Ouellette, G. P. (2008). Relations among the frequency of shared reading and 4-year-old children's vocabulary, morphological and syntax comprehension, and narrative skills. *Early Education and Development, 19*, 27– 44. https://doi.org/10.1080/10409280701838710

Shell, D. F., Murphy, C. C., & Bruning, R. H. (1989). Self-Efficacy and outcome expectancy mechanisms in reading and writing achievement. *Journal of Educational Psychology*, *81*(1), 91–100. https://doi.org/10.1037/0022-0663.81.1.91

Siegel, L. S. (1994). Working memory and reading: A life-span perspective. *International Journal of Behavioral Development, 17*(1), 109–

124. https://doi.org/10.1177/016502549401700107

Singer, L. M., & Alexander, P. A. (2017a). Reading on paper and digitally: What the past decades of empirical research reveal. *Review of Educational Research, 87*(6), 1007–1041. https://doi.org/10.3102/0034654317722961

Singer, L. M., & Alexander, P. A. (2017b). Reading across mediums: Effects of reading digital and print texts on comprehension and calibration. *The Journal of Experimental Education, 85*(1), 155–172. https://doi.org/10.3102/0034654317722961

Singer Trakhman, L. M., Alexander, P. A., & Berkowitz, L. E. (2017c). Effects of processing time on comprehension and calibration in print and digital mediums. *The Journal of Experimental Education, 87*(1), 101–115.

Singer Trakhman, L. M., Alexander, P. A., & Silverman, A. B. (2018). Profiling reading in print and digital mediums. *Learning and Instruction, 57*, 5–17. https://doi.org/10.1080/00220973.2017.1411877

Singh, A., & Alexander, P. A. (2023). *Reading and digital listening habits of college students*. Manuscript in preparation.

Singh, A., & Alexander, P. A. (2022). Audiobooks, print, and comprehension: What we know and what we need to know. *Educational Psychology Review*, 1–39. https://doi.org/10.1007/s10648-021-09653-2

Smagorinsky, P. (1989). The reliability and validity of protocol analysis. *Written Communication, 6*(4), 463–479. https://doi.org/10.1177/0741088389006004003

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360 –

407. https://doi.org/10.1598/RRQ.21.4.1

Stanovich, K. E., West, R. F., & Harrison, M. R. (1995). Knowledge growth and maintenance across the life span: The role of print exposure. *Developmental Psychology, 31*, 811– 826. https://doi.org/10.1037/0012- 1649.31.5.811

Sticht, T.G. (1974). Auding and reading: A developmental model. https://doi.org/10.1037/e457482004-001

Støle, H., Mangen, A., & Schwippert, K. (2020). Assessing children's reading comprehension on paper and screen: A mode-effect study. *Computers & Education*, *151*, 1–13 . https://doi.org/10.1016/j.compedu.2020.103861

Stone, N. J. (2000). Exploring the relationship between calibration and self–regulated learning. *Educational Psychology Review*, *12*(4), 437–475. https://doi.org/10.1023/a:1009084430926

Swanson, E., Vaughn, S., Wanzek, J., Petscher, Y., Heckert, J., Cavanaugh, C., ... & Tackett, K. (2011). A synthesis of read-aloud interventions on early reading outcomes among preschool through third graders at risk for reading difficulties. *Journal of Learning Disabilities, 44*(3), 258–275. https://doi.org/10.1177/0022219410378444

Sundre, D. L., & Moore, D. L. (2002). The Student Opinion Scale: A measure of examinee motivation. *Assessment Update, 14*(1), 8–9.

Surber, J. R., & Schroeder, M. (2007). Effect of prior domain knowledge and headings on processing of informative text. *Contemporary Educational Psychology, 32*(3), 485–498. https://doi.org/10.1016/j.cedpsych.2006.08.002

Spivey, M. J., & Geng, J. J. (2001). Oculomotor mechanisms triggered by imagery and memory: Spontaneous eye movements to objects that aren't there. *Psychological Research, 65*, 235–241. https://doi.org/10.1007/s004260100059

Tein, J. Y., Coxe, S., & Cham, H. (2013). Statistical power to detect the correct number of classes in latent profile analysis. *Structural Equation Modeling: A Multidisciplinary Journal, 20*(4), 640–657. https://doi.org/10.1080/10705511.2013.824781

Touchstone Applied Science Associates. (1983). *Degrees of reading power*. New York: Author.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*(2), 207–232. https://doi.org/10.1016/0010-0285(73)90033-9

van de Vijver, F. J. R., & Harsveld, M. (1994). The incomplete equivalence of the paper-and-pencil and computerized versions of the General Aptitude Test Battery. *Journal of Applied Psychology, 79*(6), 852–859. https://doi.org/10.1037/0021-9010.79.6.852

van den Broek, P., Young, M., Tzeng, Y., & Linderholm, T. (1999). The Landscape model of reading: Inferences and the online construction of memory representation. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 71–98). Lawrence Erlbaum Associates Publishers.

van Dijk, T. A., & Kintsch, W. (1982). *Strategies of discourse comprehension*. Academic Press.

van Gog, T., Paas, F., van Merriënboer, J. J. G., & Witte, P. (2005). Uncovering the problem-solving process: Cued retrospective reporting versus concurrent and retrospective reporting. *Journal of Experimental Psychology: Applied, 11*(4), 237–244. https://doi.org/10.1037/1076-898X.11.4.237

Verlaan, W., Pearce, D. L., & Zeng, G. (2017). Revisiting Sticht: The changing nature of the relationship between listening comprehension and reading comprehension among upper elementary and middle school students over the last 50 years. *Literacy Research and Instruction*, *56*(2), 176–197. https://doi.org/10.1080/19388071.2016.1275070

Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. In J. Hagenaars, & A. McCutcheon (Eds.), *Applied latent class analysis*. (pp. 89–106). Cambridge University Press.

Vermunt, J. D., & Vermetten, Y. J. (2004). Patterns in student learning: Relationships between learning strategies, conceptions of learning, and learning orientations. *Educational Psychology Review, 16*(4), 359–384. https://doi.org/10.1007/s10648-004-0005-y

Vidal-Abarca, E., Mañá, A., & Gil, L. (2010). Individual differences for self-regulating task-oriented reading activities. *Journal of Educational Psychology, 102*(4), 817–826. https://doi.org/10.1037/a0020062

Waddell, D. L., & Blankenship, J. C. (1994). Answer changing: A meta-analysis of the prevalence and patterns. The *Journal of Continuing Education in Nursing, 25*(4), 155–158.

Webb, S., & Chang, A. C. (2015). Second language vocabulary learning through extensive reading with audio support: How do frequency and distribution of occurrence affect learning?. *Language Teaching Research*, *19*(6), 667–686. https://doi.org/10.1177/1362168814559800

Wigfield, A. and Cambria, J. (2010). Expectancy-value theory: Retrospective and prospective. In T.C. Urdan & , S.A. Karabenick (Eds.) *The decade ahead: Theoretical perspectives on motivation and achievement* (Vol. 16, pp. 35–70). Emerald Group Publishing. https://doi.org/10.1108/S0749-7423(2010)000016A005

Wigfield, A., Guthrie, J. T., Perencevich, K. C., Taboada, A., Klauda, S. L., McRae, A., & Barbosa, P. (2008). Role of reading engagement in mediating effects of reading comprehension instruction on reading outcomes. *Psychology in the Schools, 45*(5), 432–445. https://doi.org/10.1002/pits.20307

Wiley, J., & Voss, J. F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology, 91*(2), 301–311. https://doi.org/10.1037/0022-0663.91.2.301

Wolfe, J. H. (1970). Pattern clustering by multivariate mixture analysis. *Multivariate Behavioral Research, 5*(3), 329–350.

Wirtz, D., Kruger, J., Miller, D. T., & Mathur, P. (2010). On first versus false instincts. In R. M. Arkin, K. C. Oleson, & P. J. Carroll (Eds.), *Handbook of the uncertain self* (pp. 160–175). Psychology Press.

Wise, S. L., & DeMars, C. E. (2005). Low Examinee Effort in Low-Stakes Assessment: Problems and Potential Solutions. *Educational Assessment, 10*(1), 1–17. https://doi.org/10.1207/s15326977ea1001_1

Wise, S. L., & Gao, L. (2017). A general approach to measuring test-taking effort on computer-based tests. *Applied Measurement in Education, 30*(4), 343–354. https://doi.org/10.1080/08957347.2017.1353992

Wittkower, D.E. (2011). A provisional phenomenology of the audiobook. In M. Rubery (Ed.), *Audiobooks, literature, and sound studies* (Vol. 12, pp. 216–231). Routledge.

Wolf, M. (2018). *Reader, come home: The reading brain in a digital world*. Harper.

Wolf, M. C., Muijselaar, M. M. L., Boonstra, A. M., & de Bree, E. H. (2019). The relationship between reading and listening comprehension: shared and modality-specific components. *Reading and Writing*, *32*(7), 1747–1767. https://doi.org/10.1007/s11145-018-9924-8

Zhang, X. (2013). Foreign language listening anxiety and listening performance: Conceptualizations and causal relationships. *System, 41*(1), 164–177.

Zimmerman, B. J. (2000). Attaining Self-Regulation: A social cognitive perspective. In M. Boekaerts, P.R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). Academic Press. https://doi.org/10.1016/B978-012109890-2/50031-7

Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation: Where metacognition and motivation intersect. In D.J. Hacker, J. Dunlosky, & A.C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 311–328). Routledge.