Exploring Automatic Speech Recognition as a Writing Tool

Danial Mehdipour-Kolour

A Thesis

In the Department

of

Education

Presented in Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy (Education) at

Concordia University

Montreal, Quebec, Canada

December 2023

CONCORDIA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Danial Mehdipour-Kolour

Entitled: Exploring Automatic Speech Recognition as a Writing Tool

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy Education

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

Dr. Jennifer Burton Chair

Dr. Walcir Cardoso Thesis Supervisor

Dr. Sarita Kennedy Examiner

Dr. Julie Corrigan Examiner

Dr. Paul John Examiner

Dr. Shannon McCrocklin External Examiner

Approved by _____
Dr. Sandra Martin-Chang, Graduate Program Director

_____
Dr. Pascale Sicotte, Dean of Arts and Science

December 2023

Abstract

Exploring Automatic Speech Recognition as a Writing Tool

Danial Mehdipour-Kolour, Ph.D.

Concordia University, 2023

The aim of this dissertation is to explore the use of automatic speech recognition (ASR) for writing purposes and to assess its pedagogical effectiveness as a writing tool. Considering that ASR is useful for enhancing second language (L2) oral skills and pronunciation (Liakin et al., 2015), it is worth examining ASR's pedagogical features and limitations; whether ASR can be accepted as a writing tool; and how ASR-assisted writing can contribute to the development of writing skills. This dissertation begins with an overview of the ASR technology and its architecture, as well as the technology's advancements throughout time, followed by an analysis of the components of writing skill, transcription modes, and related challenges. The remainder comprises three interconnected manuscript-based chapters (Chapters 2 to 4, respectively), each of which addresses a distinct aspect of three overarching research questions: (a) What are the pedagogical benefits and limitations of ASR use for L2 pedagogy and for writing purposes? (Manuscript A); (b) What are learners' perceptions of ASR use as a writing tool? (Manuscript B); and (c) What behaviors do learners employ when engaged in ASR-assisted writing? (Manuscript C).

The first manuscript systematically reviews the pedagogical potential and limitations of using the ASR technology in L2 learning and, more particularly, in L2 English writing. The second manuscript explores writers' perceptions of employing ASR for writing to reveal whether writers accept ASR as a tool for composing texts. This manuscript also compares

native and nonnative English writers' perceptions in terms of using ASR for writing. Finally, the third manuscript examines the behaviors that writers employ when composing texts with ASR and whether there is a relationship between the choice of behavior utilized and the lexical diversity of an ASR-written text.

This dissertation will further our understanding of users' perceptions and acceptance of ASR as a writing tool by identifying both the potentials and limitations of the technology. Given that very little research has been conducted to investigate the use of ASR systems for L2 writing by adult writers (as far as the researcher is aware, there is only one study in this field, and the handful of studies available focuses on writing by native speakers or on writers with learning disabilities), this dissertation will open a new line of inquiry in the field of L2 writing development.

Acknowledgements

Throughout my academic journey, there have been individuals whose steadfast support and guidance have profoundly influenced my path. At the forefront of this journey stands my esteemed supervisor, Prof. Walcir Cardoso. His sincere support for my success stands as a powerful source of motivation, highlighting the profound influence a mentor can exert on a student's life.

Furthermore, I would like to extend my appreciation to Dr. Julie Corrigan for her support and invaluable guidance throughout my academic journey. Serving as both a Research Assistant and Teaching Assistant under Julie's mentorship has been an enriching experience. Her expertise and dedication have been instrumental in shaping my academic and research pursuits. I am also particularly grateful for Dr. Sarita Kennedy thoughtful questions and constructive comments, which have consistently challenged me to think critically and strive for excellence. Sarita's perspectives on my work have not only been a source of inspiration but also a cornerstone of my academic growth. I would also like to express my gratitude to the two internal/external examiners, Dr. Paul John and Dr. Shannon McCrocklin, for their insightful comments and suggestions for improvement. Additionally, I want to thank the Department of Education for creating a supportive academic environment. The resources and opportunities provided have been essential to my growth. I'm grateful to the faculty and staff for their dedication in fostering a positive learning and research atmosphere.

Finally, I am deeply grateful for the unwavering support and understanding of my family, especially my parents. I also want to express my heartfelt thanks to my beloved Roxana Nezami, whose constant support and love have been an inspiration. I am also thankful for the support and camaraderie of my friends, Dr. Reza Hajirpasand and Arvin Shadravan, whose encouragement and shared experiences have added warmth and friendship to my academic journey.

Contribution of Authors

This manuscript-based dissertation consists of three manuscripts, as well as a general introduction (Chapter 1) that explains the research motivation and objectives, along with a concluding chapter (Chapter 5). The one *in-press* article (Manuscript A, Chapter 2), one *revised-and-resubmitted* article (Manuscript B, Chapter 3), and one *under-review* article (Manuscript C, Chapter 4) will be presented in their entirety, as stated in Concordia' 2020 edition for the Guidelines for Thesis Preparation, Examination Procedures and Regulations (p. 10-11). Given the nature of this format, it should be noted that there is some content overlap across manuscripts.

This manuscript-based dissertation was conceptualized over the course of my studies, tutorials, and meetings with my supervisor, Dr. Walcir Cardoso, and in the CALL research group meetings in the Department of Education at Concordia University. Serving as the first author for the three manuscripts included, I was the *primary* contributor to the manuscripts' conception, content, structure, data collection, and write-up, under the supervision of Dr. Walcir Cardoso. My status as the primary author in all three manuscripts (A, B and C), listed below, reflects my contribution to the research included in my dissertation:

(A) Mehdipour-Kolour, D. & Cardoso, W. (in press). A Systematic Literature Review of Automatic Speech Recognition in L2 Learning: A Case for L2 Writing. In M. Peterson & N. Jabbari (eds.), *Routledge Frontiers in computer assisted language learning*.

(B) Mehdipour-Kolour, D. & Cardoso, W. (revised, resubmitted). Users' Perceptions of ASR as a Writing Tool: An Analysis using the Technology Acceptance Model. *Computer Assisted Language Learning*.

(C) Mehdipour-Kolour, D. & Cardoso, W. (submitted). Exploring Second Language Writers' Usage Behavior of Automatic Speech Recognition: A Focus on Lexical Diversity in Narrative Texts. *Language Learning & Technology*.

Table of Contents

List of Figures

List of Tables

Chapter one: General Introduction

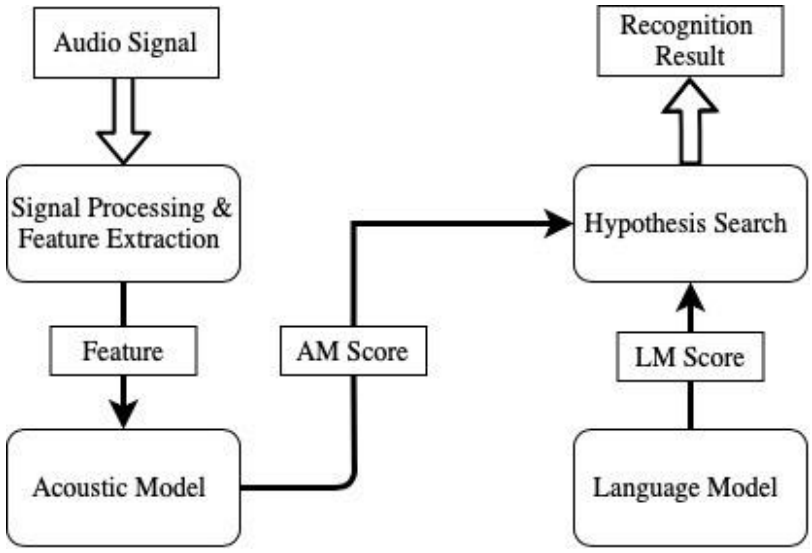
In today's world, technology has become an integral part of our daily lives, permeating every aspect of our existence. We now enjoy the convenience of utilizing various technological tools (e.g., computers, smartphones, specialized software) to fulfill our diverse needs. These tools have significantly enhanced our lives by providing efficiency, flexibility, and instant access to information. One noteworthy example of such a tool is automatic speech recognition software (ASR), which has revolutionized the way humans interact with computers (Enge, 2020). ASR introduces a novel *input medium* that facilitates seamless interaction between humans and machines (Enge, 2020).

Introduction to ASR

ASR, a piece of software built into many digital devices (e.g., smartphones, computers), recognizes spoken words by decoding the acoustic input and produces output in the form of text. ASR entails a set of procedures that are widely categorized into four approaches: Signal processing and feature extraction, acoustic model, language model, and hypothesis search using artificial intelligence (Yu & Deng, 2014). Figure 1.1 depicts the typical procedures of ASR systems.

Figure 1.1

Procedures of ASR Systems (Yu & Deng, 2014)



The *signal processing and feature extraction* receives an audio signal as input, improves it by eliminating noise distortions, transforms the audio signal from temporal-domain (i.e., the analysis of speech signals in the time dimension by processing the speech waveform as a sequence of intensity levels over time) to frequency-domain (i.e., the analysis of speech signals in the frequency dimension by decomposing the speech waveform into its constituent frequency components), and extracts prominent features suited for acoustic models. The *acoustic model* combines data of phonetics with acoustic properties, uses (as input) the prominent features extracted by the feature extraction approach, and provides an acoustic model (AM) score. The *language model* approximates the likelihood of a predicted word order by identifying the relationship between words from training data; the result of this approximation is a language model (LM) score. Finally, the *hypothesis search* approach aggregates AM and LM scores based on the predicted word order and produces the word order with the greatest score, which functions as the recognition result (Yu & Deng, 2014).

ASR Systems: History and Development

Since the 1950s, numerous ASR systems have been developed, showcasing the continuous advancements in this technology. Notably, more recent examples include *Dragon Naturally Speaking*, *Siri*, and *Google Voice Typing* (the technology adopted in the current dissertation). The earliest ASR system, the *Audrey System*, was developed in 1952 by Bell Laboratories; it was designed to recognize numbers so it could facilitate handsfree calling, a useful affordance while driving (Daniels & Iwago, 2017). By the 1960s, ASR systems like IBM's Shoebox advanced from recognizing numbers to words, which was originally limited to 16 English words that increased to 1,000 words in the 1970s, but with a large number of recognition errors (Cardoso, 2022). ASR's accuracy rate has grown from around 80% in the late 1990s and early 2000s to 95% (and above) in recent years, and to 99% with well-trained systems adapted for particular professions such as medicine (Leijten & Van Waes, 2006). The advancement of ASR extends far beyond mere improvements in accuracy and the number of spoken words the system can transcribe. Indeed, advances in artificial intelligence have empowered ASR to move beyond the limitations of recognizing only discrete words, enabling it to now adeptly recognize and understand more natural speech patterns.

Three distinct categories of ASR have been established, each classified according to the type of spoken input they can handle: (1) discrete word recognition; (2) connected word recognition; and (3) continuous speech recognition (Jurafsky & Martin, 2000; Liakin et al., 2015). Early ASR systems were considered *discrete word recognition* that could only recognize words articulated separately. Advanced ASR systems at the time upgraded to *connected word recognition*, meaning that the user had to insert a short pause after a group of connected words was spoken. Currently, ASR systems can recognize *continuous speech*, meaning that the user can produce words/phrases or entire sentences without the necessity of inserting pauses between words/phrases or after sentences.

Thanks to advancements in speech recognition algorithms, three distinct types of ASR systems have emerged, distinguished by the level of prior training necessary for optimal functionality: (1) speaker-dependent; (2) speaker adaptive; and (3) speaker-independent (Liakin et al., 2015; Rosen & Yampolsky, 2000; Young & Mihailidis, 2010). Being among the world's earliest ASR systems, *speaker-dependent* ASR requires user-specific training data, meaning that the user must train the system with his/her own speech and, as a result, ASR can function effectively for that user only. Considered a more contemporary ASR system, *speaker adaptive* ASR is gradually tailored to the acoustic features of a particular user's speech utilizing a small sample of speeches, with no prior user-specific training required. Lastly, as the most advanced ASR system, *speaker-independent* ASR functions effectively independent of user-specific training data, meaning that the system does not require speech samples from a particular user because it has been already trained with speech samples collected from a variety of users. It is important to note that, much like other technologies, speech recognition software is still evolving. Therefore, further advancement of ASR will impact related research, particularly on using the technology as a writing tool and its effects on the writing process. At this point, the technology has indeed introduced a new mode for composing text that provides new insights into components of the writing process (Leijten & Van Waes, 2006).

ASR and Writing

With hindsight, the dominance of 'silent' reading and writing appears to be being challenged by a 'secondary orality' assisted by numerous technologies (Honeycutt, 2004). As Baron (2000) argues, "In the nineteenth century, mass schooling reinforced the growth of popular literacy … and the more oral, accessible style of middle-class novels and short stories became the norm. These trends have continued ever since" (p. 77). Even though the shift from silent to voiced writing has fundamentally changed the writer-text relationship, the

implementation of ASR as a writing tool can affect this relationship yet again because the technology combines the oral and written attributes of the text (Leijten & Van Waes, 2006).

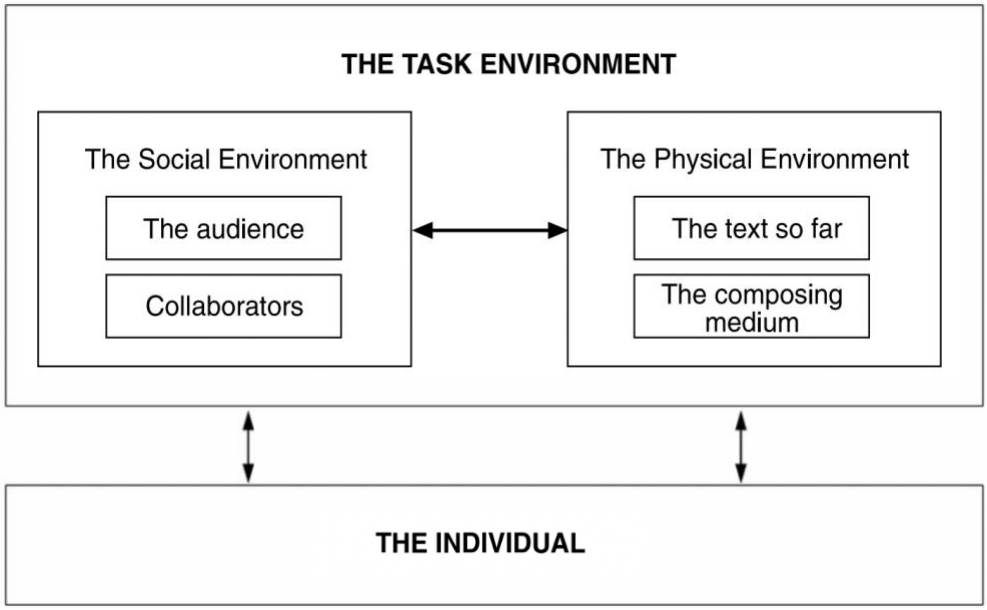
As a hybrid mode of writing, ASR combines the advantages of typing in word processors and traditional dictating via recorders. Writing using ASR has three unique features closely connected to the writer's interaction with the 'text produced so far' (see Chapter 2 for further discussion): Visibility, immediacy, and spelling accuracy. ASR *immediately* creates a *visual* representation of the text produced, which is *accurate* in terms of spelling. In comparison with ASR, traditional transcription tools such as dictation provide no immediate visual representation of the text produced so far and the writer has limited opportunities for revisions during the writing process; almost entire revisions occur mentally prior to dictating the text to the recorder. It can thus be argued that texts produced by traditional dictation are highly linear (Leijten & Van Waes, 2005). It is important to mention that the transcribed output of dictation can enhance opportunities for revisions, thereby increasing the non-linearity of the text produced so far.

ASR, on the other hand, allows writers to continuously plan, review, and revise the text produced so far at all phases of the writing process. According to the Hayes' (1996) writing model, the writing process consists of two major components: the task environment and the individual (see Figure 1.2). The task environment comprises the social environment and the physical environment. The social environment deals with the audience (i.e., the intended group of people to which a text is written) and with any collaborators involved in the writing process. As Hayes' model shows, the physical environment incorporates two essential parts: the *text produced so far*, which affects and determines the writer's continuing efforts, thus playing an important role in managing the writing process (Leijten & Van Waes, 2006); and the *composing medium* (e.g., keyboarding in a word processor, writing with ASR), which influences stages of the writing process, such as planning and revising (Haas,

1987; Weigle, 2010). In terms of composing media, different ASR engines have been developed (e.g., Google Voice typing, Microsoft Dictate, Apple Voice Control, Amazon Transcribe). In the next section, Google Voice Typing, the speech recognition engine of Google (and the ASR adopted in this research) will be presented.

Figure 1.2

Hayes' model of writing (adapted from Hayes, 1996)



Example of an ASR System: Google Voice Typing

Google Voice Typing is a seamlessly integrated feature within Google Docs, making it user-friendly to activate and utilize. The user only needs to open a Google Docs page, select 'Tools', and then select 'Voice typing' in the menu (see Figure 1.3). After that, a small dialogue box with a microphone icon inside appears on the left side of the page. Activating the system is as simple as clicking on the microphone icon, while deactivation can be achieved either by clicking the same icon once more or by issuing a voice command like "stop listening". As shown in Figure 1.4, there is a drop-down menu on the top of the window that allows the user to select a preferred language to interact with ASR (in this dissertation, the English language was selected). The user must ensure that the built-in (or external)

microphone of their computer works properly. After setting up the system, the user can speak their text and have ASR type it out for them (see Figure 1.5). It is worthwhile noting that, at the time of writing this dissertation, Google Voice Typing works best with the Chrome browser; it is either not available on some browsers (e.g., Safari) or not fully compatible with other browsers (e.g., Microsoft Edge).

Figure 1.3

Activation of Google Voice Typing in Google Docs

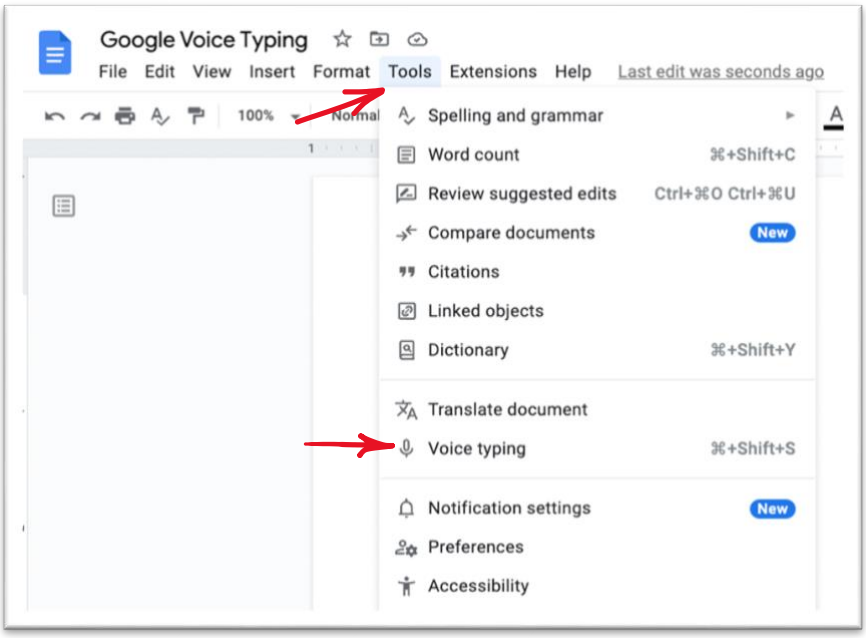


Figure 1.4

Selecting the language

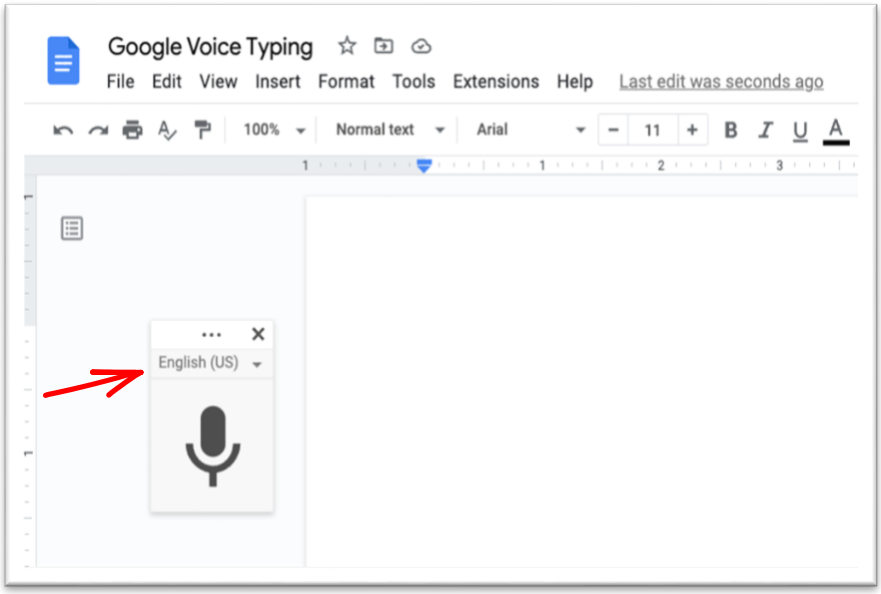
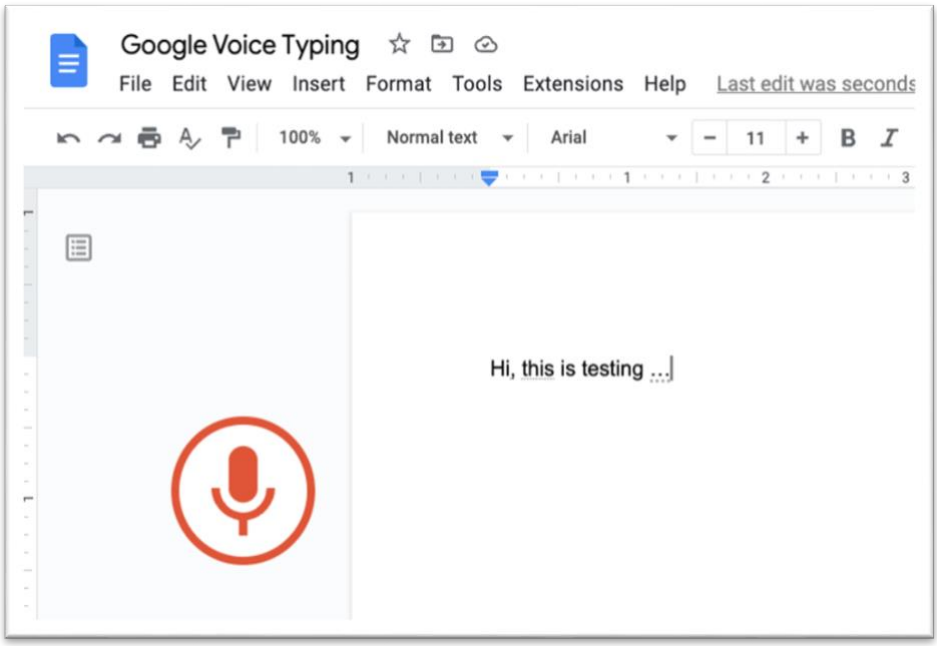


Figure 1.5

Speaking the text to ASR



Google Voice Typing allows the writer to give voice commands to punctuate the text (e.g., by saying 'comma'). Further, using voice, the writer can edit the text by the command 'select [part of the text]' followed by the command 'delete'. This feature can reduce the writer's use of keyboard and mouse. The additional benefit of the system is that the spoken texts are saved in the Google cloud, making it simple for learners to share and access them from various devices.

Why Google Voice Typing: Technical and Theoretical Motivation

The present dissertation adopts Google Voice Typing (the Google's ASR system) for a number of reasons. First, Google Voice Typing is free to use (and freely available to embed in other software); it is also one of the most readily accessible ASR systems (Daniels & Iwago, 2017). The system is available in Google Docs, which is linked to the Google cloud, thus making it easy for students to access the system through any smartphone to engage in ASR-assisted writing anytime, anywhere. Secondly, Google Voice Typing is highly accurate

(95% recognition accuracy) even to foreign-accented speech (McCrocklin & Edalatishams, 2020). Thirdly, since Google Voice Typing is available in more than 100 languages, it can encourage L2 learners to write with ASR using their native or other languages. Lastly, Google Voice Typing (and ASR systems in general) fulfil Chapelle's (2001) criteria for adopting computer-assisted language learning (CALL) tools, which will be discussed below.

Chapelle (2001) proposes six criteria to determine if a technological tool is appropriate for use in an L2 learning context: Language learning potential, learner fit, meaning focus, authenticity, positive impact, and practicality. *Language learning potential* refers to the extent to which there is potential for constructive form-focused attention. *Learner fit* is described as the extent of opportunities for language engagement under appropriate situations, considering learner characteristics (e.g., proficiency level, age). *Meaning focus* denotes the degree of learners' focus on language meaning. *Authenticity* refers to how closely the CALL activity mirrors interesting activities that learners may engage in outside of the classroom context to improve their target language skills. *Positive impact* is specified as the benefits of participating in the CALL activity. Finally, *practicality* is described as the availability of sufficient resources to facilitate the use of the CALL activity.

To direct a 'judgmental evaluation' of a CALL activity, Chapelle (2001) proposes the following questions, guided by the abovementioned criteria:

(1) Language learning potential: Do activity situations provide ample opportunity for a constructive focus on form?

(2) Learner fit: Is the complexity level of the intended linguistic forms appropriate for learners to advance their language skills? Does the activity fit the targeted learners' characteristics?

(3) Meaning focus: Is learners' attention focused mainly on language meaning?

(4) Authenticity: Does the CALL activity closely mirror language activities that students may engage in outside of the classroom context? Will learners be able to understand the link between the CALL activity and activities performed outside of the classroom context?

(5) Positive impact: Will the use of the activity help learners gain knowledge of the target language and language learning strategies? Will the activity use be conducive to useful L2 pedagogical practices? Will the use of the activity lead to a positive learning experience for both learners and teachers?

(6) Practicality: Are the available technology and resources adequate to make the CALL task successful?

ASR systems such as Google Voice Typing meet these criteria. Regarding *language learning potential*, ASR has been shown to be an effective tool in promoting input exposure and language production (Liakin et al., 2017), and form-focused instruction (Jiang et al., 2021; Young & Wang, 2014). For instance, focusing on the phonological aspect of language (i.e., an example of a language form), Liakin et al.'s (2017) study demonstrates the language learning potential of ASR for teaching L2 French at the segmental (vowel /y/) and suprasegmental levels (liaison). In another study, Bodnar et al. (2017) finds that ASR-enabled CALL systems can provide beneficial conditions for practicing syntactic forms orally where immediate feedback on oral outputs does not evoke negative affect in learners.

In terms of *learner fit*, ASR is adaptable to learners' interests, needs, and characteristics since the technology provides a personalized human-computer interaction enabling a wide array of learners (e.g., of different ages and proficiency levels: Young & Wang, 2014; of different learning styles: Liakin et al., 2015) to benefit from the use of the technology in their own learning.

As for *meaning focus*, ASR can promote human-machine interactions that direct learners' focus toward meaning. For example, Walker et al. (2011) incorporates ASR technology to foster meaning-focused interactions. In their study, they stimulate interviews (authentic interactions) between a nurse and a patient, and ASR elicits meaningful responses from participants, thus having the potential to draw learners' focus to meaning (see Chapter 2 for a discussion of how ASR fosters meaning negotiation).

Concerning *authenticity*, ASR use can mirror language activities that learners find interesting to engage in outside of the classroom (Nguyen et al., 2018). In the same context, Yu et al. (2016) use ASR technology to provide an authentic practice setting where learners can interact with their teachers, classmates, and content (e.g., through real-life interactions) to improve their spoken English.

Regarding *positive impact*, research has found that the implementation of ASR systems results in positive L2 learning experiences, beneficial L2 pedagogical practices, and effective learning about the target language and language learning strategies (Mroz, 2018; Van Lieshout & Cardoso, 2022).

Lastly, concerning *practicality*, ASR systems are readily available and easy to use anytime anywhere: they only require an internet-connected device (e.g., smartphone, laptop) and the use of a natural human ability, one's voice (van Doremalen et al., 2016).

ASR and CALL Research

It is important to examine the pedagogical appropriateness of ASR systems for writing purposes, especially L2 writing, because very little is known about their pedagogical potential. According to Cardoso (2022), CALL researchers often follow a chronological framework for developing or exploring new technologies that have potential for pedagogical purposes. The framework involves four stages:

1) Development: at this stage, the emphasis is on the development of a new technological tool, commonly drawing on insights from the L2 acquisition, CALL, computer science. This stage is not pertinent to the present thesis since ASR systems have already been developed and are broadly used in voice assistants (e.g., Apple Siri, Google Voice Typing).

2) Exploration: this stage entails an exploration of the pedagogical affordances, potentials, as well as limitations of a *new* technological tool (i.e., as a follow-up to its development) or of an *existing* technology, such as ASR (see Manuscript A in Chapter 2 for the implementation of this type of research).

3) Assessing suitability: this stage involves an examination of learners' perceptions of using the technological tool, which can be carried out by measuring different constructs (e.g., the technology's usability, learners' attitudes towards a technology). Once a new technological tool is available and potentially appropriate for use in a CALL context, one approach a CALL researcher might undertake is to assess its suitability considering learners' perceptions via quantitative (e.g., surveys), qualitative (e.g., semi-structured interviews), or mixed methods (see Manuscript B in Chapter 3).

4) Assessing pedagogical effectiveness: the last stage involves an assessment of the technology's pedagogical effectiveness. Studies focusing on this stage investigate whether the target technological tool affects learning (e.g., by analyzing user behavior; see Manuscript C in Chapter 4).

Breakthroughs in ASR technology, along with the rising ubiquity of smartphones and personal computers (Mehdipour-Kolour et al., 2018), have enhanced the overall accessibility and efficiency of the technology (McCrocklin & Edalatishams, 2020). These features have stimulated interest in a large number of CALL researchers to *explore*, *assess the suitability* and *pedagogical effectiveness* of ASR use in L2 learning, with an emphasis on various language skills: Speaking and pronunciation (Cucchiarini et al., 2012; Liakin et al., 2015);

listening (Chen, 2011; Mirzaei et al., 2017). However, a limited number of CALL studies have probed the pedagogical affordances and limitations, and learners' perceptions of ASR use for writing with learners without writing (or learning) difficulties (for example, see MacArthur, 1999, 2000). Moreover, no study to date has examined ASR as a tool for L2 writing purposes. Therefore, this dissertation contributes to the CALL literature by investigating ASR's pedagogical suitability and potential as a tool for writing.

The Current Dissertation

Guided by Cardoso's (2022) chronological framework, the present research for a manuscript-based PhD dissertation incorporates stages 2 through 4 to examine a readily available ASR system (Google Voice Typing) as a tool for L2 writing. Despite the unavoidable overlap, each chapter concentrates on one stage specifically: chapter 2 investigates stage 2 (pedagogical affordances, potentials, and limitations of ASR in L2 pedagogy, particularly in writing), chapter 3 is informed by stage 3 (users'/learners' perceptions of using ASR for writing), while chapter 4 is directed towards stage 4 (assessing the pedagogical effectiveness of ASR in learners' use of writing strategies). Accordingly, the dissertation is divided into three major topics, each of which represented by a manuscript driven by an overarching research question that reflects stages 2 through 4 of Cardoso's framework.

1) Manuscript A (Chapter 2): What are the pedagogical benefits and limitations of ASR use for L2 pedagogy and for writing purposes?
2) Manuscript B (Chapter 3): What are learners' perceptions of ASR use as a writing tool?
3) Manuscript C (Chapter 4): What behaviors do learners employ when engaged in ASR-assisted writing?

The first manuscript (stage 2: Exploration), adopting the software evaluation framework (SEF; Burston, 2003; Hubbard, 1988), analyzes ASR technology to provide the theoretical underpinning and motivation for using ASR in a CALL context. In addition, the manuscript systematically investigates ASR's pedagogical affordances (e.g., accessibility), potential (e.g., creation of opportunities for language production), and limitations (e.g., recognition errors in texts) in the L2 learning context and as a tool for writing purposes. The second manuscript (stage 3: Assessing suitability) employs the technology acceptance model 2 (TAM2), developed by Venkatesh and Davis (2000), to investigate how writers perceive the use of ASR as a transcription mode and how native and nonnative English writers' perceptions of using this tool for writing differ. Finally, the third manuscript (stage 4: Assessing pedagogical effectiveness) explores aspects of ASR's pedagogical effectiveness by examining how learners interact with ASR, specifically analyzing the behaviors they employ when producing text (e.g., using ASR exclusively, ASR use in combination with a keyboard and mouse), and how the use of certain behaviors might affect the lexical diversity of ASR-written texts.

Chapter Two: A Systematic Literature Review of Automatic Speech Recognition in L2 Learning: A Case for L2 Writing

Automatic speech recognition (ASR), a system built into many word processors (e.g., Microsoft Word) and digital devices (e.g., smartphones, laptops), decodes any audio input to identify spoken words and generates text as output. Since the advent of speech recognition technology in the 1950s, researchers in the field of second language (L2) acquisition have discussed its merits and shortcomings (e.g., Garvin & Ladefoged, 1963; Porter, 1960). Currently, ASR has been shown to be useful for improving L2 speaking and pronunciation (Ahn & Lee, 2016; Chiu et al., 2007; Liakin et al., 2015; McCrocklin & Edalatishams, 2020), listening (Chen, 2011; Mirzaei et al., 2017), and for facilitating writing among speakers with learning difficulties (MacArthur, 1999; Quinlan, 2004)). As will be demonstrated in this literature review, while ASR has proven useful in language learning, especially in improving speaking and pronunciation skills, there is still a lack of research on the effectiveness of ASR for writing purposes.

This chapter first provides the rationale for ASR use in L2 education using the software evaluation framework (SEF; Burston, 2003; Hubbard, 1988) and insights from Second Language Acquisition (SLA; e.g., Output hypothesis – Swain, 1985; Interaction Hypothesis – Long, 1983; Negotiation of Meaning - de Vries et al., 2015; Gass, 1997). Next, following insights from Hubbard (2006) and the CALL literature (e.g., Chapelle, 2001; Egbert & Shahrokni, 2018), our review of 79 empirical studies examines the pedagogical potential and limitations of the technology use for L2 learning, particularly for writing purposes.

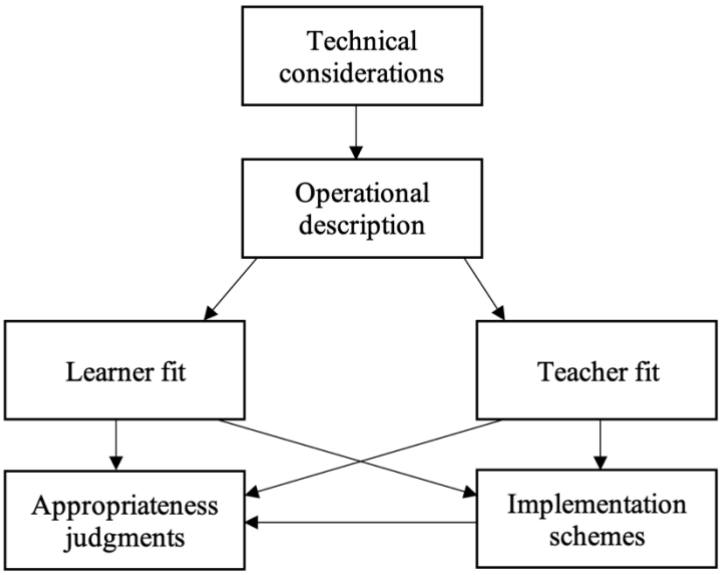
Software Evaluation Framework: ASR Use in L2 Education

This review will initially analyze ASR technology through the Software Evaluation Framework (Burston, 2003; Hubbard, 1988), one of the most prominent evaluation frameworks

in the field of CALL. Figure 2.1 demonstrates the core structure of the framework, which will be discussed next.

Figure 2.1

Software Evaluation Framework: Core Components (Burston, 2003; Hubbard, 1988)



Technical Considerations

From a technical standpoint, SEF raises the most fundamental question of whether the software under consideration will be *accessible* on the devices that students and/or teachers plan to implement in their pedagogical experience (Hubbard, 2006). ASR is a readily accessible technology because it can operate on the vast majority of desktop computers and mobile devices, regardless of operating system. Further, most smart devices have built-in ASR capabilities, so it is no longer necessary to install or purchase costly speech recognition packages such as Dragon NaturallySpeaking. In terms of system configuration, ASR does not require powerful processors and it functions well with any built-in microphone found in computers and mobile devices.

Operational Description

The operational description component of SEF provides a review of the elements of the system and how they function independently or are managed by the user (Hubbard, 2006). In this review the operational description includes the *presentational schemes* only, which will be explained in the context of ASR.

According to Hubbard (2006), presentational schemes are delineated by some subcategories, which in the present review includes the screen layout or interface, timing, and user input. The s*creen layout*/*interface* is about the basic features related to the look and feel of the system (e.g., fonts, quality of graphics). In general, ASR systems have simple and intuitive interfaces with a microphone icon that is activated/deactivated with a click or tap.

Timing (or responsiveness) refers to the time it takes for content material or feedback to appear on the screen. With new ASR systems, the user does not have to pause speech after uttering each word or phrase in order for the system to properly recognize the user output; rather, the user can output continuous speech and receive immediate and fairly accurate feedback in the form of text on the screen (Saadatzi et al., 2017).

User input is concerned with how the user interacts with the system (e.g., clicking, speaking, typing) in response to prompts appearing on the screen. As ASR systems have a voice user interface, the user input is provided in the form of speech. Using speech, a natural means of interaction, enhances the ease of use and interactivity of ASR (van Doremalen et al., 2016), because it is easier and more natural than using a keyboard and mouse (Selouani et al., 2008), particularly for children or users with mobility impairment and/or neurological disorders such as dysgraphia (Cavalier & Ferretti, 1996), and likely for L2 learners.

Teacher Fit

Hubbard (2006) asserts that, in the context of computer-assisted language teaching (CALT), teacher fit represents the evaluator's assumptions about: (a) the nature of language

(i.e., the interrelationship between language, communication, and culture); (b) how languages are learned; and (c) the capabilities of CALL as a facilitator and delivery tool for language learning (Egbert & Shahrokni, 2018). Together, these assumptions develop the evaluator's *CALT-based approach*, which provides grounds for creating *approach-based evaluation criteria*.

The CALT-based approach that supports the pedagogical use of ASR derives from the Output Hypothesis (Swain, 1985) and the Interaction Hypothesis (Long, 1983). The Output Hypothesis stresses that language output is essential for L2 learning, as it requires learners to communicate the desired meaning using the target language, thus helping them to notice the gap in their interlanguage (Swain, 2005). Based on the Interaction Hypothesis, *interaction* and *feedback* (i.e., for negotiation of meaning) play a major role in L2 learning (de Vries et al., 2015; S. Gass, 1997). An example involving human-machine interaction followed by feedback and negotiation of meaning (see forthcoming discussion) would be when an ASR user intends to produce the following output: "when online, I'm bombarded by pictures, news", but the system incorrectly shows something like "online karma I'm bombarded by pictures comma news". These types of errors can occur because ASR systems often require the user to verbally specify punctuation marks. This can result in incorrect transcriptions, as the system may misinterpret the spoken punctuation ("karma" instead of *comma*) or fail to recognize it altogether ("comma" as a word, not a punctuation mark).

Based on the abovementioned ASR output, the *user* engages in meaning negotiation in the form of repeating, correcting, or rephrasing the underlined words above to ensure the system can correctly process the intended message. It is important to note that, in contrast to meaning negotiation involving two people, in which both sides are usually involved in the process, the negotiation of meaning in this context happens solely on the user's side (*intra-interaction* or within a learner's mind, using Chapelle's 2003 terms), and the ASR acts as a

facilitator of meaning negotiation. Motivated by the output hypothesis and the interaction hypothesis, ASR creates opportunities for L2 learners to produce language and receive immediate feedback (via faultless orthography) on their output, thus fostering a sense of interaction with the computer (Cardoso, 2022; Chapelle & Jamieson, 2008; de Vries et al., 2015).

Learner Fit

According to SEF, learner fit consists of two sub-variables: *learner variables* and *syllabus*. While the former includes variables such as age, proficiency level, and learner needs, the latter includes items such as linguistic objectives (e.g., pronunciation accuracy) and language skills (e.g., writing).

With respect to learner variables, research shows ASR's capability to accommodate different factors: age (Wang & Young, 2015); native language (de Vries et al., 2015); proficiency level (Liakin et al., 2015; Mroz, 2018); learner needs and interests (Chapelle & Jamieson, 2008; Liakin et al., 2015); and learning style (Hsu, 2016). Concerning syllabus, the available research points to the pedagogical benefits of ASR in teaching pronunciation (Liakin et al., 2017; Sidgi & Shaari, 2017), speaking (Ahn & Lee, 2016), and listening (Mirzaei et al., 2017). However, only a handful of studies have shown the benefits of ASR for adult writers (Leijten & Van Waes, 2005; Wetzel, 1996).

In this review, learner fit includes two other components related to both learner variables and syllabus: *language difficulty* and *program difficulty* (Hubbard, 2006). Language difficulty describes the level of linguistically related complications caused by a number of factors (e.g., grammar complexity, clarity of pronunciation in audio-visual materials). Program difficulty, on the other hand, is concerned with how complicated it is to work with the system.

Research indicates that the ASR-based issues related to language difficulty involve recognition errors that can occur, for example, when a user with non-native accented speech interacts with an ASR system (Liakin et al., 2017; McCrocklin, 2019). Although ASR systems such as Google's are highly accurate even to non-native accented speakers (McCrocklin, 2019), their recognition accuracy can be affected by factors such as pronunciation that is unintelligible to the system. In terms of program difficulty, studies show that learners find it easy to interact with ASR (van Doremalen et al., 2016).

Appropriateness Judgments

As Hubbard (2006) posits, considerations related to teacher fit and learner fit together form judgments of the appropriateness of a CALL tool for a specific context. The fulfillment of teacher and learner fit affects the outcome of the user's *appropriateness judgment* of whether to adopt (i.e., accept, integrate, and use new technology), reject, or sustainably use the CALL tool under consideration, ASR in our case. It is essential that the appropriateness judgment also affect *how* and *when* to use the system.

Implementation Schemes

Finally, according to Hubbard (1996), *learner training* can contribute to the successful implementation of CALL tools. With ASR, learners need training on technical (e.g., system setup, voice commands) and pedagogical issues (e.g., to interpret orthographic feedback), thus learning to interact with the system effectively to fulfill their pedagogical needs.

Summary and the Present Study

As indicated above, the pedagogical use of ASR fulfills most of the criteria set forth by Hubbard's (1988) and Burston's (2003) Software Evaluation Framework and, more importantly, it has the potential of addressing some of the attributes that promote L2 acquisition (e.g., ASR provides learners with ample opportunities to produce output, interact

with the tool, and negotiate for meaning). However, it is still unclear what the literature, as a whole, reports about the benefits and limitations of the technology.

The goal of this study is to systematically investigate the pedagogical benefits and limitations of using ASR in L2 education and for the development of L2 English writing, considering research published between 1990 and 2021 (see forthcoming discussion for the rationale). As Grant and Booth (2009) describe, "systematic reviews seek to draw together all known knowledge on a topic area […] incorporating quantitative, qualitative and mixed method studies" (p. 102). To review and analyze the relevant ASR literature, as discussed in the previous sections, this study adopts the following well-recognized CALL/SLA theories: output hypothesis (Swain, 1985), learner autonomy (Cardoso, 2022; Chapelle, 2001), interaction hypothesis (Long, 1983), developmental model of writing (Berninger & Winn, 2006), meaning negotiation (Gass, 1997), cognitive load theory (Sweller, 1988), and learner fit (Hubbard, 2006).

To fulfil its goals, the study addressed the following question:

- What does the empirical literature on ASR use in L2 education reveal about the potential benefits and limitations of the technology in terms of
    a. L2 education in general; and
    b. L2 writing specifically?

Method

Inclusion Criteria

The inclusion criteria were restricted to empirical studies, and full-text journal articles published in English. Dissertations, theoretical papers, short papers, and reports were excluded because they are not usually peer-reviewed and might fail to meet the standards of a high-quality research study. The search aimed at two independent ASR uses: (1) as a tool for L2 educational purposes in studies published between 2012 and 2021; and (2) as a writing

tool (including the use of transcription and dictation) between 1990 and 2021. Regarding L2 education, the 2012 and 2021 period was chosen because it is within the recent 10 years that free commercial ASR systems have been equipped with artificial intelligence technology and machine learning algorithms that have boosted ASR systems accuracy and responsiveness. Concerning the use of ASR as a writing tool, the 1990 and 2021 period was chosen because ASR research on writing, though limited in quantity, was initially conducted in the 1990s.

The search was accomplished through multiple databases including ERIC, Google Scholar, ScienceDirect, JSTOR, and SAGE, using the following keywords: *automatic speech recognition, computer-assisted pronunciation training, automated speaking assessment, second language acquisition, computer-/mobile-assisted language learning, writing, writing process, transcription modes, writing development,* and *automatic captioning.*

The initial search generated 714 entries related to ASR in L2 education and 73 entries related to ASR as a writing tool, as shown in Figures 2.2 and 2.3 respectively. After pre-screening, 197 and 32 entries respectively were considered for potential inclusion in the analysis. Following the criteria established for this study, this resulted in 97 journal articles focused on ASR in L2 education and 19 articles focused on ASR as a writing tool. Afterward, titles and abstracts were screened, which led us to remove 33 and 4 articles respectively from the initial search results because they: (1) did not focus on L2 skills; (2) centred on describing the architecture of ASR engines; (3) targeted speech pathology; (4) did not use ASR for transcribing speech (in the case of articles involving "writing"); or (5) were non-empirical and non-experimental in nature. There remained 64 articles on the use of ASR in L2 education (Figure 2.2), and 15 articles on ASR-based writing (Figure 2.3).

These data were imported to Dedoose (version 8.0.35), a web software for analyzing mixed methods data, to conduct thematic analyses. Adopted from the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA; Page et al., 2021), Figures 2.2

and 2.3 demonstrate how potential publications moved through the different stages of the present systematic review.

Figure 2.2

Summary of systematic search and review: ASR in L2 education

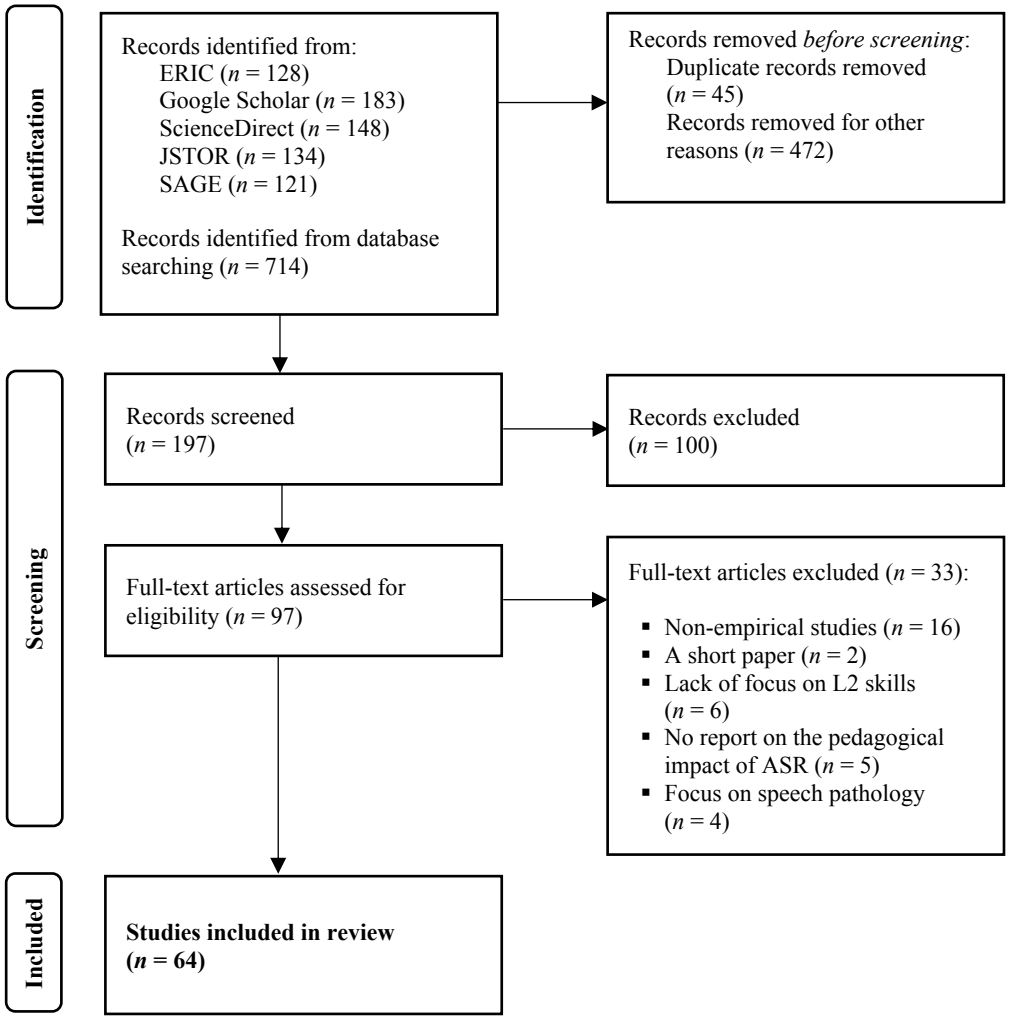
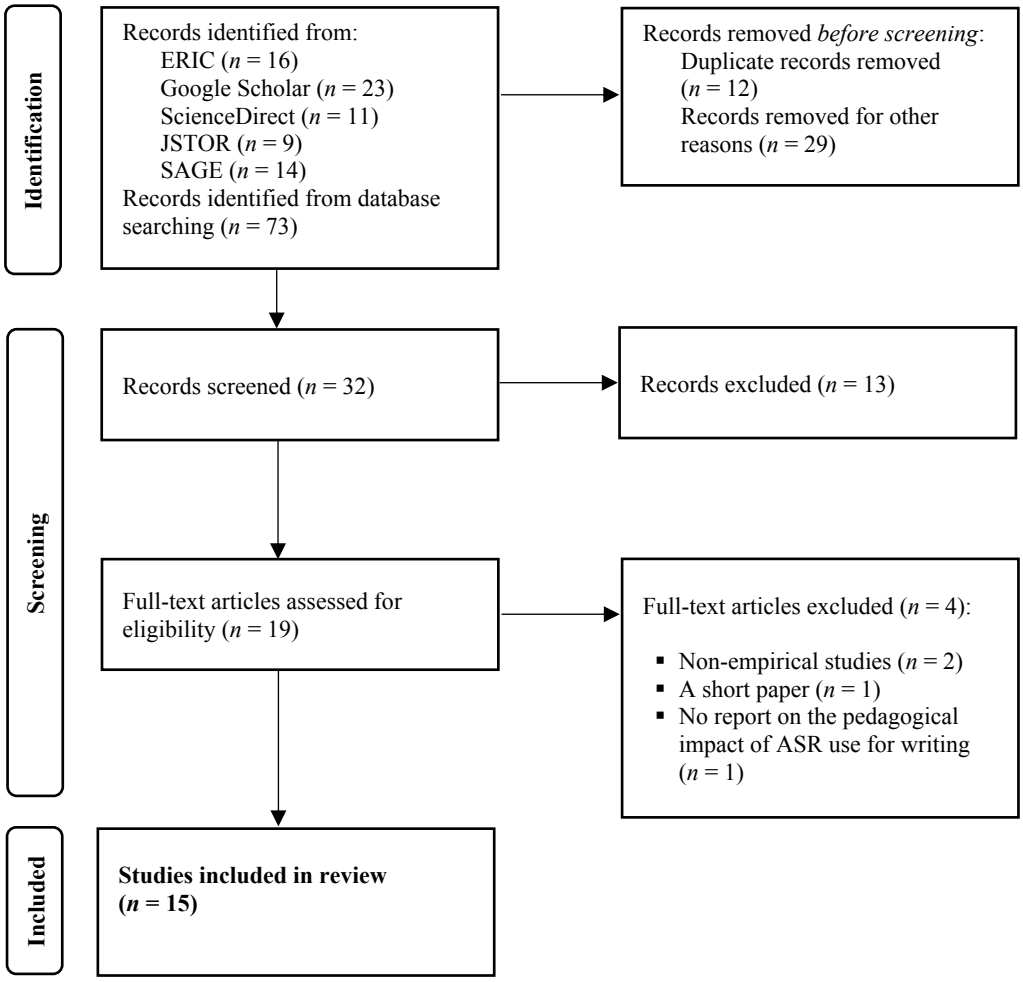


Figure 2.3

Summary of systematic search and review: ASR as a Writing Tool



Coding Procedures

The selected studies were coded based on a number of variables, starting with the pedagogical affordances and limitations of the ASR technology. Coding targets also included L2 learning skills (e.g., emphasis on improving pronunciation), L2 teaching practices (e.g., utilizing multimedia to make learning more engaging), writing development (e.g., enhancing writing fluency), effects of transcription methods (e.g., handwriting, voice typing) on writing quality, and the motivation of the study. The pedagogical affordances of ASR were coded based on CALL-oriented SLA principles such as the tool's ability to provide opportunities for interaction, exposure to language, learner motivation, authenticity, and autonomous learning

(Chapelle, 2001; Egbert & Shahrokni, 2018). L2 learning and teaching practices were coded based on the output hypothesis (Swain, 1985), interaction hypothesis (Long, 1983), and feedback (Chapelle & Jamieson, 2008). Coding of writing skills and transcription modes was based on the developmental model of writing (Berninger & Winn, 2006), which postulates that three interrelated component skills are engaged during the process of ASR writing: *transcription skills*, *oral language skills*, and *executive functions* – due to the nature of this review, only transcription skills were considered in the coding process.

The variables were first coded through *a priori* codes and then through *inductive* codes. As Johnson and Christensen (2019) describe, "*a priori* codes or preexisting codes … [are] developed before or at the very beginning of the current research study [and] inductive codes … are generated by the researcher by directly examining the data during the coding process" (p. 548). Coding was mutually inclusive, meaning that each article could be coded for several features of pedagogical affordances and limitations of ASR, L2 learning, writing skills and transcription modes, and the theoretical motivation. A code was assigned based on the existence of specified and recurrent themes or patterns found in the data. This coding procedure involved systematically identifying and categorizing information that was pertinent to the research questions of the study (e.g., advantages of utilizing ASR for enhancing pronunciation skills). It is important to highlight that the assignment of codes was not directly related to the design or results of the selected studies, but rather to the content and structure of the article, and to the identification and interpretation of relevant concepts that were contained therein.

To ensure inter-coder reliability, twenty percent of randomly chosen articles (n=13 for ASR in L2 education, n=3 for ASR in L2 English writing) were independently coded by two graduate students hired for the task. The coders were familiar with SEF (the framework adopted) and the theories covered in the study; they were also trained on how to use Dedoose

to identify themes related to *a priori* codes (e.g., language exposure) and inductive codes (e.g., writing fluency). 94% of coding agreement was achieved, which is deemed acceptable (Lavrakas, 2008).

Results

To answer the research question (What does the empirical literature on ASR in L2 education reveal about the potential benefits and limitations of ASR in terms of (a) L2 education; and (b) L2 English writing?), the thematic analysis conducted found repeated themes concerning the potential benefits and limitations of using ASR in L2 education as well as L2 English writing. The themes along with the related empirical studies are reviewed below.

ASR in L2 Education: Potential Benefits

Pronunciation Development

According to the literature reviewed, 38% (*n*=24) of the studies indicate that ASR is a tool that can improve the development of L2 pronunciation, as it provides L2 learners with ample opportunities for pronunciation practice (Dillon & Wells, 2021; McCrocklin & Edalatishams, 2020; Pérez Castillejo, 2021), consequently improving their pronunciation skills in the target language (Evers & Chen, 2021; Jiang et al., 2021; Mroz, 2018). For instance, (Liakin et al., 2015) examined the production and perception of the French vowel /y/ in an ASR-based mobile learning context. Their study showed that the ASR's inherent ability to promote form-focus learning (Dekeyser, 1993) and to provide input enhancement via orthography (Chapelle, 2001), had a positive effect on French L2 learners' production of /y/. In another study, Yu et al. (2016) developed an ASR-enhanced prototype for Chinese-speaking learners of English. The ASR component of their prototype together with other feedback tools (e.g., audio replaying) provided the learners with authentic pronunciation practice, which resulted in improvements in their English pronunciation (see also Sidgi and

Shaari (2017) for similar results involving Arabic-speaking English learners who utilized EyeSpeak, an ASR-supported system).

Opportunities for L2 Output

Lack of opportunities for output practice in the target language is believed to hinder language learning (Swain, 2005), particularly in *foreign* language contexts where learners have limited opportunities to interact with native (or fluent) speakers. In this review, 33% (*n*=21) of the studies show that ASR can increase opportunities for L2 output (Liakin et al., 2017; Young & Wang, 2014) via repeated practice (Richards, 2015;Daniels & Iwago, 2017) or other activities that require the recycling of newly acquired forms over time (e.g., via spaced repetition; (Chukharev-Hudilainen & Klepikova, 2016). For instance, de Vries et al. (2015) designed an ASR-based CALL system, GREET, which exposes L2 learners to spoken grammar practice to improve their speaking. The ASR component of their system allows L2 learners to interact with GREET via speech and receive immediate corrective feedback on grammatical accuracy. As per their findings, while the corrective feedback did not significantly affect grammar development, the ASR component of the system fostered interaction, and thereby enhanced opportunities for L2 practice.

Timesaving

Our analysis shows that 12.5% (*n*=8) of the studies analyzed indicate that ASR is a timesaving tool, particularly for teaching/learning pronunciation, one of the most challenging subskills to learn (Cucchiarini et al., 2012; Jiang et al., 2021; Yu et al., 2016) because it requires repeated practice and personalized feedback (Cucchiarini et al., 2012). In addition, teachers might not provide pronunciation assessment to every L2 learner, especially in a large class (Daniels & Iwago, 2017). Since ASR provides L2 learners with independent pronunciation practice and immediate feedback, it frees up classroom time (Elimat & AbuSeileek, 2014) and, accordingly, its use can encourage teachers to assign “anytime-

anywhere" pronunciation practice as homework (as the procedure was conceptualized in Liakin et al., 2015) so that class time can be devoted to communicative tasks.

Cost-effectiveness

Among the literature reviewed, a small percentage (*n*=4, 6.5%) considers cloud-based ASR systems such as Google Voice Typing as a cost-effective alternative in comparison with human assessors. For instance, ASR can be used in L2 pronunciation assessment (Cucchiarini et al., 2012; Liakin et al., 2015; McCrocklin & Edalatishams, 2020), as an accommodation in written tests to include students with disabilities in assessment systems (MacArthur & Cavalier, 2004), and in the assessment of speaking proficiency (Cox & Davies, 2012; Silva et al., 2021).

Autonomous Learning

The automatic feedback generated by ASR (via orthography) can reduce the learner's reliance on the teacher's (or other students') feedback, and thereby help them develop language skills autonomously (Bodnar et al., 2017b; Mroz, 2018). According to our findings, 16% (*n*=10) of the studies reviewed show that the interaction that L2 learners have with ASR promotes autonomous learning of speaking and pronunciation (Cardoso, 2022; McCrocklin, 2016). Research indicates that a pedagogically important aspect of ASR-based interaction is that the L2 learner can use the tool's feedback to notice and identify possible errors in pronunciation (Mroz, 2018). As Elimat and AbuSeileek (2014) found in their study, this interaction increases learners' autonomy, consequently impacting their pronunciation performance more positively than practice in pairs or groups.

Explicit Teaching

Among the studies examined, 5% (*n*=3) demonstrate that ASR can provide language learners with explicit pronunciation practice (Elimat & AbuSeileek, 2014; Liakin et al., 2015; Mroz, 2018). Specifically, L2 English learners who receive ASR-assisted explicit teaching of

pronunciation, such as spelling-to-sound patterns, have been shown to develop better grapho-phonemic knowledge (Evers & Chen, 2021).

Inclusive Learning Environment

An inclusive learning environment is one in which all learners, irrespective of L1s, educational level, or preferred learning styles, feel valued and are afforded a sense of belonging. One of the challenges that L2 teachers face is how to handle the diversity of learners within the classroom (Young & Wang, 2014), including different L1s, ages, proficiency levels, and learning styles. To tackle this challenge, the implementation of ASR can be useful in that it provides individualized human-computer interaction that allows all learners to benefit from the technology (de Vries et al., 2015). In this review, 6.5% (*n*=4) of the literature discusses the inclusivity attribute of ASR, particularly in terms of recognizing speakers of different L1s (de Vries et al., 2015), accommodating a variety of learning styles (Liakin et al., 2015), as well as proficiency levels and age groups (Young & Wang, 2014).

In conclusion, our literature review reveals that the pedagogical use of ASR is commonly associated with benefits such as improving pronunciation (found in 35% of all studies consulted) and providing learners with opportunities for L2 output practice (33%). While the benefits of ASR for language learning are numerous and varied, our findings also suggest that its use is timesaving (12.5%) and cost-effective (6.5%), and that it creates opportunities for autonomous learning (16%) and explicit teaching (5%), in a learning environment that is inclusive (6.5%). Overall, the use of ASR holds potential in facilitating language learning and teaching, particularly in terms of pronunciation improvement.

ASR in L2 Education: Potential Limitations

According to nearly 33% (*n*=21) of the literature reviewed, one of the common limitations reported is the low *recognition accuracy* of the system, particularly for accented non-native speakers (Daniels & Iwago, 2017). 12.5% (*n*=8) of the studies analyzed report L2

learners' frustration caused by the poor recognition of ASR (McCrocklin, 2019; McCrocklin & Edalatishams, 2020) and by ASR's inability to produce explicit corrective feedback (Liakin et al., 2017). Additionally, despite the human-like interaction that ASR affords (Cardoso, 2022), about 5% (*n*=3) of the studies examined reveals that ASR-based interaction lacks some features of effective oral communication such as nonverbal cues (e.g., Mroz, 2018).

To conclude, while a considerable number of studies demonstrate that ASR is a useful tool for L2 education, particularly pronunciation (e.g., Liakin et al., 2015; Mroz, 2018), very few studies have investigated the use of ASR as a writing tool. Indeed, the vast majority of ASR-based writing research has only focused on the use of ASR for writers with learning disabilities (Fox et al., 2021; MacArthur, 2000; Quinlan, 2004). Below we delve into the emerging field of ASR-based L2 writing, which remains relatively unexplored and underrepresented in the current literature.

ASR in English L2 Writing

Our search yielded only 15 studies that examine ASR as a writing tool. Within these studies, the majority involve native speakers of English (Hartley et al., 2003; Leijten & Van Waes, 2005; Reece & Cummings, 1996) or focus on using ASR to facilitate writing among people with learning disabilities (Cavalier & Ferretti, 1996; Fox et al., 2021; Quinlan, 2004).

ASR in English L2 Writing: Potential Benefits

The reviewed literature suggests that there are three main benefits of using ASR for writing, as will be discussed below.

Process Approach to Writing. 40% (*n*=6) of the studies analyzed reveal that one of the distinctive benefits of ASR is its ability to promote a process approach to writing via the production of voice-inputted text (VIT) on the screen. VIT refers to the developing text that appears on the screen immediately after the user inputs *speech* into ASR. VIT creates a significant degree of *non-linearity* in writing, as it allows the writer to constantly read,

review, plan, and revise the text throughout the writing process (Haas, 1989; Leijten & Van Waes, 2005; Wood et al., 2002). In this context, (Breetvelt et al., 1996) found that reviewing and revising VIT considerably improved lawyers' writing quality. The significance of VIT is reflected in models of writing development, as they emphasize the importance of reviewing, planning, and revising the developing text for proficient writing (Flower & Hayes, 1980; Hayes, 1996; Kellogg, 1996). In contrast, traditional dictation practices limit writers to linearly produce text, as they do not provide an opportunity to see, plan, and revise the text dynamically. This means that writers must make revisions mentally before dictation (Leijten & Van Waes, 2006).

In an experiment conducted by Reece and Cummings (1996), children were asked to write a text using three different transcription modes: handwriting, dictation, and ASR. In the dictation mode, texts were dictated to a typist and were free of recognition errors, whereas the ASR-written texts included recognition errors. The researchers found that ASR writers had long pauses during which they engaged in planning their texts, and that they engaged more actively in the process of writing (e.g., through *planning*, an activity that is not observed with dictation). These findings show that ASR-written texts have higher quality than texts written by the other two modes of transcription. According to Reece and Cummings (1996), with ASR, the writer does not experience a heavily occupied working memory to maintain "some form of ongoing representation of the composition. Instead, writers can turn their attention to a consideration of higher-level aspects of the composition task" (p. 375).

Speech-based Input Tool. 67% (n=10) of the literature examined shows that ASR offers a speech-based *input tool* (Leijten & Van Waes, 2006), reducing the writer's dependence on a keyboard or mouse (Saadatzi et al., 2017) – the remaining 10 studies did not directly explore ASR as a speech input medium; instead, they focused on analyzing the written output generated through ASR technology or on other affordances of ASR use for

writing. As a "more natural" and faster input mode (speech is a feature that most humans possess), speaking-to-write facilitates the writer's ability to produce texts (Quinlan, 2004). This is an important benefit, particularly for L2 writers who have limited L2 knowledge and hence a stronger need to focus more on language than content (Weigle, 2010).

Speech-based input tools can provide L2 writers with an alternative tool that can facilitate *fluency* in writing, as they can help these writers to overcome handwriting- and/or typing-related issues (MacArthur & Cavalier, 2004) such as poor spelling or forgetting the layout of keyboard letters, because they are different from the order of the alphabet. The use of ASR minimizes such transcription-related issues (MacArthur, 1999), frees up working memory resources, thus reducing cognitive load. As a result, the writer can primarily focus on text production, which is necessary for writing development (McCutchen et al., 1994).

Inclusive Writing Environment. 20% (*n*=3) of the reviewed studies indicate that ASR creates an *inclusive* writing environment where writers with learning disabilities, such as dysgraphia, can produce text. Additionally, ASR allows writers to adopt their own writing behaviors such as pausing (period of inactivity during writing that allows writers to engage in planning and decision making; Matsuhashi, 1982) to compose texts (Hartley et al., 2003; Leijten & Van Waes, 2005). For example, when writing using ASR, writers can maintain their planning and reviewing behavior, and repeatedly reread the developing text on the screen.

To summarize, our findings indicate that the most cited advantage of ASR is its ability to function as a speech-based input tool (*n*=10/15, 67%), which facilitates a process-oriented approach to writing (*n*=6/15, 40%). Additionally, ASR enables an inclusive writing environment (*n*=3/15, 20%) by allowing individuals with diverse abilities to participate in the writing process.

ASR in English L2 Writing: Potential Limitations

The literature reports a number of limitations regarding the use of ASR for writing. A small number of studies (13%, *n*=2) have considered the tool's spelling accuracy a potential limitation. Although the text produced by ASR is often free from spelling errors (Honeycutt, 2003), users' reliance on this level of accuracy can cause them to overlook recognition errors in their texts. For example, there might be cases where ASR misrecognizes a phrase like 'the use of a fluent mode' as 'the use of affluent mode'. These recognition errors (i.e., the transcription of *a fluent* as *affluent*) are not part of the writer's intended text, but due to the orthographic and phonological similarities of the two phrases, the writer might fail to realize the discrepancy. This can slow down the writing process because the writer has to be overly observant to avoid such errors (Leijten & Van Waes, 2005).

Likewise, some studies (27%, *n*=4) maintain that, although ASR-based writing eliminates spelling mistakes (e.g., in comparison with handwriting or typing), it requires a new type of error correction in texts dictated using ASR (MacArthur & Cavalier, 2004). With ASR systems, for instance, the writer must attend to misrecognized words and make corrections either by voice or a keyboard and mouse. Along the same line, MacArthur (2000) considers the speaker's intelligibility as a limitation of ASR-based writing, as the writer must pronounce words clearly and accurately to minimize ASR's misrecognition.

Correcting misrecognized words is of paramount importance, because as Wetzel (1996) stated, "too much meaning is lost when one of every four words is substantially incorrect. If the errors are to be corrected later, too much is left to revise" (p. 377). Likewise, Leijten and Van Waes (2005) pointed out that erroneous words on the screen can be distracting and make it more cognitively demanding for the writer to create a coherent text. Further, monitoring and correcting incorrect words can be challenging for L2 writers with poor literacy skills (MacArthur, 1999).

Furthermore, some studies (20%, *n*=3) have viewed text formatting as a potential limitation of ASR-based writing and have posited that dictating non-text elements can impose on the writer additional cognitive burdens (MacArthur & Cavalier, 2004). For instance, ASR requires the writer to dictate punctuation marks (e.g., period, comma) and give voice commands (e.g., [new line], [select word], [delete word]) to format the text. In terms of using voice commands, there can be two challenges to tackle: (a) the writer needs to remember the necessary commands (Wetzel, 1996); and (b) voice commands can be mistakenly recognized as text, rather than commands.

Discussion and Concluding Remarks

This study reviewed the literature published between 1990 and 2021 to answer the question of what the empirical literature on ASR in L2 education reveals about its potential benefits and limitations when used in (a) L2 education and (b) L2 writing. Many benefits were found regarding ASR use in L2 education; these include: the promotion of *opportunities for pronunciation development* (38%, *n*=24), increased *opportunities for L2 output* (33%, *n*=21), incentive for *autonomous learning* (16%, *n*=10), and the tool's efficiency as *timesaving* (12.5%, *n*=8). Our research has also found that ASR can facilitate *explicit teaching* (*n*=3, 5%) by, for instance, creating opportunities for students to obtain personalized feedback. Additionally, ASR can foster *an inclusive learning environment* (*n*=4, 6.5%) by accommodating diverse learning abilities and styles. Finally, ASR can be a *cost-effective* tool for assessing pronunciation (*n*=4, 6.5%), providing an efficient and reliable alternative to human evaluators. However, as pointed out in a number of studies, the use of ASR in L2 education is not free from limitations. These include (a) low recognition accuracy of the system to non-native accented speech; (b) lack of explicit corrective feedback; (c) relatively long time required for the correction of recognition errors; and (d) lack of communicative features such as nonverbal cues.

Concerning the potential benefits of ASR in L2 English writing, three benefits were frequently referenced: ASR's ability to offer a *speech-based writing tool* 67% (*n*=10), to promote a *process-oriented approach to writing* (40%, *n*=6), and to establish an *inclusive writing environment* (20%, *n*=3). However, as indicated in the review, ASR in L2 writing presents some potential limitations to writers, which include (a) the excessive effort to revise misrecognized words in the VIT (the text produced); (b) the complexity of text formatting (e.g., dictation of non-text elements and voice commands such as 'new line', 'comma' – for further details on voice commands, see Google Docs Editors Help Center/Type with your voice); and (c) the writer's speaking intelligibility while writing using ASR.

This review has not only examined the motivation for using ASR in L2 education, but it has also discussed why and how it can benefit L2 learners. According to Macaro et al. (2012), not only should CALL research examine whether a technology tool is useful, but it should also examine the reason for its usefulness and its contribution to the theory in the field. Accordingly, this review explicitly provides an analysis of the pedagogical benefits and limitations of ASR use in general L2 pedagogy and in L2 writing.

Based on our findings, future studies could investigate (a) the affordances of ASR systems as a transcription mode for English L2 writers without learning disabilities; (b) native and non-native English writers' perceptions of using ASR for writing purposes; and (c) the impact of ASR as a writing tool on the quality of texts of different genres (e.g., narrative, argumentative). In terms of text quality, further research is needed to specifically explore the linguistic features that predict ASR-written text quality. (D. McNamara et al., 2010) described these features as coreference and connectives, representing *cohesion*, and syntactic complexity, lexical diversity, and word characteristics, representing *language sophistication*. Future research could also be devoted to the investigation of these linguistic features in ASR-

written versus typed texts to show if and how the mode of transcription affects the features that predict text quality.

Chapter Three: Users' Perceptions of ASR as a Writing Tool: An Analysis using the Technology Acceptance Model

Speaking is the most common mode of communication across cultures: it is simpler, faster, and more efficient than writing (Selouani et al., 2008). Human speech emerged long before writing, meaning that early humans were able to communicate through speech sounds before they started to communicate via text. Although there is no consensus on when speech was developed, research indicates that the development of human speech dates back to 60-70 thousand years ago (Perreault & Mathew, 2012). However, the history of human's ability to communicate via *writing* is remarkably more recent (dating back to around six thousand years ago; (Powell, 2009). Since the earliest form of writing, which was invented in Sumer near the Persian Gulf, writing tools have gone through many technological evolutions (Bolter, 2001): from Sumerians' tablets and styluses around 3400 BC, the printing press in the 1500s, commercial typewriters in 1874, ballpoint pens in 1888, computer keyboards in 1964, to thumb keyboards on smart devices (e.g., smartphones) in the late 1990s.

The present study using a survey method investigates ASR, an innovative technology that can be characterized as a writing tool. ASR, a built-in system of most current technological devices (e.g., desktop computers, smartphones), recognizes spoken words (i.e., voice input) and converts them into text (i.e., text output), which is simultaneously produced on the screen. In a large-scale study, Enge (2020) surveyed more than 1000 people and found that about 55% of the surveyees used voice searches and voice commands via ASR on their smartphones. This demonstrates that people are becoming increasingly accustomed to using voice to interact with their smart devices, particularly after the advent of digital personal assistants such as Google Assistant and Amazon Alexa (Moussalli & Cardoso, 2019).

While the use of voice as an *input medium* is gradually on the rise (Enge, 2020; Smith et al., 2021), it has remained unknown how users perceive ASR as a tool for writing. Thus,

this study focuses on users' perceptions because, as Cardoso (2022) maintains, determining users' perceptions is one of the steps to find out whether they will accept a new technology, after it becomes available. In addition, understanding users' perceptions of technology can greatly impact the success of its implementation. For instance, teachers who are less confident with technology are less likely to integrate it into their teaching (Ertmer et al., 2012). No study, to the knowledge of the researchers, has examined the perceptions of ASR users for writing. In addition, there is no previous evidence showing how native and nonnative English writers perceive the use of ASR as a transcription mode. To rectify these gaps in the literature, this study adopted the Technology Acceptance Model 2 (TAM2; Venkatesh & Davis, 2000)—an extension of the technology acceptance model (TAM; Davis, 1989)—to investigate (1) how writers perceive the use of ASR as a transcription mode, and (2) how native and nonnative English writers' perceptions of using this tool for writing differ.

Background

Importance of Writing

Writing is an essential aspect of the development of communicative competence and literacy (Behizadeh & Engelhard, 2011). However, no one is born a competent writer. As Darwin (1871) explained, "man has an instinctive tendency to speak, as we see in the babble of our children, whereas no child has an instinctive tendency to […] write" (p. 55). While effective writing is becoming increasingly prominent to address academic, occupational, and personal needs (Weigle, 2010), research has shown a declining proficiency in learners' writing abilities (Annamalai & Tan, 2015; Healey & Gardner, 2021)

This issue can be closely related to how effectively the component skills employed in the writing process function. Based on the Developmental Model of Writing (Berninger & Winn, 2006), a written text is the result of three interconnected component skills engaged in working memory during the writing process: *transcription skills*, *oral language skills*, and

executive functions. This model, which identifies the constraints that affect the process of writing, can be used for exploring learners' writing development (Drijbooms et al., 2017), as will be discussed below. It should be noted that the present study focuses on transcription skills only; as such, it excludes oral language skills and executive functions from its scope.

Transcription Modes and ASR-based Writing

Transcription refers to "the translating of … language representations in working memory into written symbols on the printed page" (Berninger et al., 1996, p. 197). It should be noted that transcription (of any mode) is an essential aspect of the writing process as it enables writers to record their thoughts and ideas. Transcription modes can be divided into four types: *handwriting* via pen and paper, *typing* via a keyboard*, dictating* via a recorder (with no text produced on the screen), and *voice typing* via an ASR system with *voice-inputted text* produced on the screen.

Each of these transcription modes is characterized by several writing mechanics. As described by MacArthur (1999), writing mechanics refer to "all the transcription processes involved in getting words into print—handwriting or typing, spelling, capitalization and punctuation, and formatting" (p. 170). Lack of competence in the mechanics of writing inhibits the writer's ability to produce ideas, arrange ideas coherently, and assess and revise sentences properly, thereby causing negative impact on the length and quality of text (MacArthur, 1999). Besides, different writing mechanics occupy working memory (i.e., the short-term information storage of human mind; Miller et al., 1960) to a varying degree, thus affecting the writer's cognitive load differently (for further details, see Cognitive Load Theory (CLT) in Sweller, 1988).

For instance, handwriting can overload the writer's working memory, i.e., increasing their cognitive load (in comparison with typing) because the writer must remember orthographic representations, separate them into graphemes, remember the relevant

graphemic forms, and utilize graphomotor or handwriting skills to produce text (Quinlan, 2006). Evidence shows that poor performance in either of the aforesaid stages can negatively impact vocabulary retrieval, and thereby impede the writer's ability to effectively turn ideas into words (Bourdin, 2000). However, *fluent* handwriting results in reduced cognitive load, which enhances the writer's efficiency in other important writing processes, e.g., idea generation (Bourdin & Fayol, 1994). In contrast to handwriting, typing can occupy the writer's cognitive load to a lesser extent in that it requires simpler graphomotor skills and eliminates spelling barriers, especially while using word processors, thus fostering fluency in transcription (Quinlan, 2006). As research shows, fluent transcription is necessary for writing development (Kellogg, 2001; McCutchen, 1996).

Dictation can be conducive to reducing cognitive load (in comparison with handwriting and typing) because it removes problems concerning the mechanics of writing (e.g., spelling, slow typing of low-frequency words), thus providing a useful transcription mode for novice writers (Berninger et al., 1996; MacArthur & Cavalier, 2004). However, "standard" dictation devices do not produce text output on the screen, so the writer loses the chance of reviewing what has been written so far to see if it makes sense to the reader, to remember what ideas he/she has developed and how they are organized, to think about what needs detailed explaining, to develop further ideas, and to plan the remainder of the text (MacArthur, 1999). Research indicates that the absence of text output on the screen negatively affects the components involved in the writing process (e.g., planning, editing), and thereby leads to a final text that lacks appropriate organization, coherence, clarity, and focus (Graham, 1990; Reece & Cummings, 1996).

Interestingly, ASR offers a hybrid mode of transcription featuring the characteristics of both dictation and typing. It is comparable to dictation because the writer must use voice to input text. It is also comparable to typing because ASR reproduces the developing text on

the screen. Indeed, speech, a natural human capability in comparison with keyboarding (Selouani et al., 2008), turns ASR into a transcription mode that can considerably decrease the writer's cognitive load (Quinlan, 2004), possibly contributing to an improvement in writing quality (Arcon et al., 2017). ASR surpasses mere dictation and typing as it optimizes the transcription efficiency and frees up cognitive resources for higher-level writing processes (e.g., ideation, planning, revision).

Benefits of ASR Use as a Transcription Mode

People can benefit from writing using ASR in three principal ways: it promotes a process approach to writing, it establishes an inclusive writing environment, and it uses speech as an input medium.

When the user inputs speech into ASR, the system immediately produces the developing text on the screen. This affordance promotes non-linearity in writing, which refers to the idea that writing is not necessarily a sequential process in which writers start at the beginning and advance through to the final product. Rather, writing tends to be a recursive process of revision and reorganization of the text (Flower & Hayes, 1981). ASR can foster non-linear writing by allowing writers to simply move back and forth between various parts of their text to read, revise, and edit them, or add new ideas during the writing process without losing focus, thus promoting a *process approach* to writing (Leijten & Van Waes, 2005). Adopting a process approach can assist writers in becoming more creative and adaptable in their writing process, enabling them to link different elements of their text and explore and develop ideas in a more natural way. Indeed, most of the writing models underline the importance of the process approach for proficient writing (Hayes, 1996). From this perspective, Breetvelt et al. (1996) found that reviewing and revising the developing text improved participants' writing quality. Conversely, with dictation (without a simultaneous transcriber), the writer misses the chance of reviewing and revising what was written because

dictation does not produce the developing text simultaneously. This implies that the writer using dictation has no choice but to produce text linearly, a process which is deprived of dynamic planning and revisions (Leijten & Van Waes, 2006).

ASR also provides an *inclusive writing environment* where writers with and/or without learning disabilities can use the system to effectively produce text (Quinlan, 2004). Moreover, ASR does not encourage one writing behavior (e.g., planning, revising) over another. As Leijten and van Waes (2005) indicate, a shift from one transcription mode to another brings about changes in writing behavior (e.g., pausing behavior, defined as "Moments of scribal inactivity during writing [that] reflect [writers' involvement with] cognitive planning and decision-making behavior" (Matsuhashi, 1982, p. 270). Interestingly, new ASR systems respond to L2 accented speech highly accurately (McCrocklin & Edalatishams, 2020), thus likely to embrace a wider range of L2 writers.

Finally, ASR affords an *input medium* operating with *speech*, a natural and efficient mode of communication which is faster than writing (Moats & Tolman, 2009; Selouani et al., 2008). Using speech for inputting text can minimize transcription-related issues (e.g., bad handwriting, poor spelling), which are shown to hinder writing (MacArthur, 1999; MacArthur & Cavalier, 2004). Accordingly, a speech-based input medium can promote fluency in transcription, i.e., production of more content efficiently (Quinlan, 2006), an affordance that is of paramount importance for L2 writers with limited knowledge about the L2's writing system. These writers have a stronger need to focus more on language than content (Weigle, 2010), so using speech for inputting text can direct L2 writers' attention towards content production (rather than language), a requisite for writing development (McCutchen et al., 1994). It is worth noting that ASR cannot supplant human cognition in creating text; however, it can facilitate the writing process by lowering the cognitive burden

associated with transcribing, allowing writers to concentrate more on the components of the writing process, such as reviewing and revision.

However, a closer look at the literature reveals a number of gaps. For example, research investigating adult English as L1 and L2 writers' perceptions of using ASR, particularly recent ASR systems (e.g., Google Voice Typing) as modes of transcription for creating narrative texts remains extremely scarce. In addition, most of the ASR studies conducted in the field to date involve young writers (e.g., Arcon et al., 2017) or native writers with learning disabilities (e.g., MacArthur, 2000; Quinlan, 2004). Moreover, studies that compare how native and nonnative English writers perceive the use ASR as a transcription mode do not exist. One way to fill these gaps in the literature is to investigate the issue by adopting a top-down approach, using technology acceptance frameworks such as TAM2, which will be reviewed below.

Technology Acceptance Model 2

This study uses TAM2, a widely used theoretical model in the field of information systems to measure users' attitude and acceptance of new technologies. Developed by Venkatesh and Davis (2000), TAM2 is an extension of TAM, originally created by Davis (1989). As shown in Figure 3.1, TAM suggests that the user's intention to adopt a new technology is grounded on two determinants: *Perceived usefulness* and *perceived ease of use*. While perceived usefulness refers to what extent the user believes that using the technology will improve his/her performance, perceived ease of use refers to what extent the user believes that using the technology will be effortless. According to Venkatesh and Davis (2000), perceived ease of use is a stronger determinant than perceived usefulness because, when relevant technologies are equal, the user is more likely to perceive the easier-to-use technology as more useful.

The more recent version of this model (TAM2) encompasses two additional determinants that are relevant to the current study: Social determinants (subjective norm) and cognitive determinants (output quality and result demonstrability).

Subjective norm refers to the user's perception that is shaped by the influence of important people (e.g., a school board) who think the user should or should not use the new technology (Fishbein & Ajzen, 1975) . As shown in Figure 3.1, *subjective norm* directly affects the attitude and intention to use because the user may choose to adopt a new technology, though personally an unfavorable choice, if he/she believes that other important people think that the technology should be adopted (Venkatesh & Davis, 2000). Despite the direct effect of subjective norm on user intention, such effect may eventually become eroded as user experience increases (Hartwick & Barki, 1994). Novice users' knowledge and attitude about a new technology are often limited and ambiguous. Thus, they rely on the ideas of other people to form their perceptions, and hence their intentions of using or rejecting the technology. However, as TAM2 posits, increased experience with the technology will diminish the direct effect of subjective norm over time and, as a result, will lay the basis for the user's perceptions and intentions of using the technology (Venkatesh & Davis, 2000).

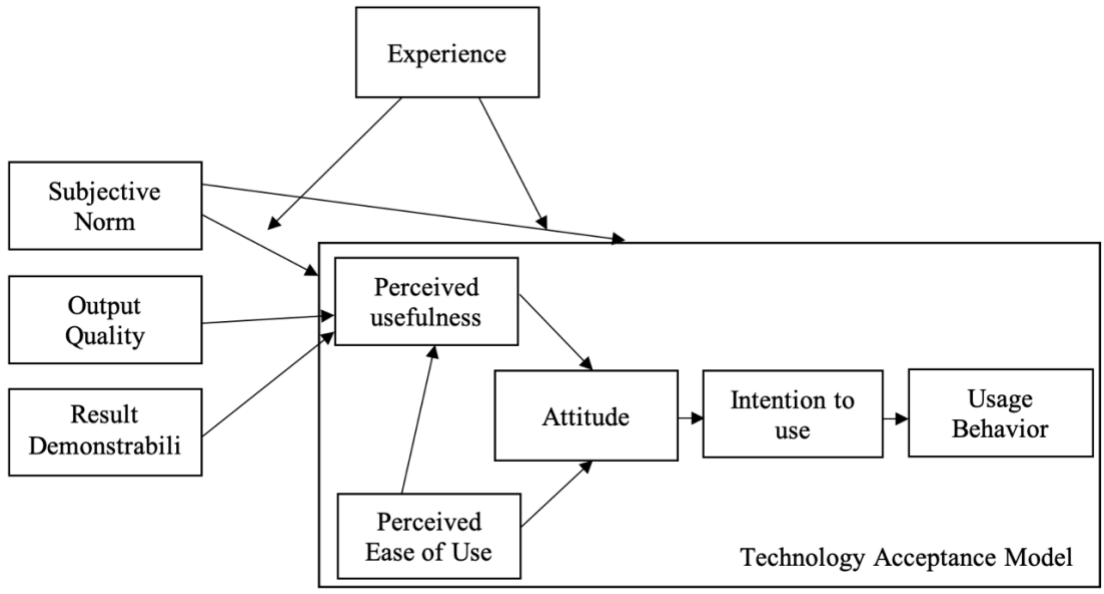
Considered a cognitive determinant, *output quality* is described as the extent to which the technology can perform the target tasks effectively. According to Davis et al. (1992), perceived output quality correlates with perceived usefulness. This implies that when a set of equally relevant technologies being available, the user is inclined to perceive the usefulness of a technological tool more positively providing that it offers the maximum output quality.

Lastly, *result demonstrability* refers to what extent the results of using the technology are tangible (Moore & Benbasat, 1991). Result demonstrability directly affects perceived usefulness, meaning that the user is more inclined to perceive the technology useful if the

correlation between technology use and positive results is evident (Venkatesh & Davis, 2000).

Figure 3.1

Adapted from TAM2 (Venkatesh & Davis, 2000)



Rationale for the Study

Despite the increasing use of voice as an *input medium* (Sohail, 2020), research in the field of voice-based writing substantially lacks investigation, particularly with new ASR systems. While studies have shown that ASR is effective for improving second language (L2) oral skills such as pronunciation (e.g., Dalby & Kewley-Port, 2013; Liakin et al., 2015) and speaking (Cox & Davies, 2012), there is limited research on the use of ASR as a transcription mode for writers without learning disabilities. Moreover, no studies have explored if native and nonnative English writers' perceptions of using ASR as a transcription mode differ. Indeed, this line of research has not been attended to in about 20 years (Leijten & Van Waes, 2005; MacArthur, 1999; Reece & Cummings, 1996), and there exists only a handful of studies on the use of ASR with young English language learners (e.g., Arcon et al., 2017) or with native writers with learning disabilities (e.g., Quinlan, 2004).

Guided by TAM2, the present study explores writers' perceptions of using ASR for writing to demonstrate whether they are willing to accept it as a transcription mode. The study also compares native and nonnative English writers' perceptions of using ASR as a writing tool. To achieve these goals, the following research questions are addressed.

RQ_1: How do writers perceive the use of ASR as a transcription mode?

RQ_2: How do native and nonnative English writers' perceptions compare in terms of using ASR as a transcription mode?

Consistent with the research questions, the following hypotheses were formulated.

H_0: Writers' perception average equals the neutral mean value of 3 (on a 1-5 scale).

H_1: Native English speakers will be more likely to perceive ASR positively than non-native English speakers.

Method

Participants

This research employed a quasi-experimental design, incorporating a one-sample t-test, and involved the recruitment of 68 participants, aged between 20 and 55. Participants were selected using a purposive sampling method and subsequently divided into two groups: Native speakers (NSs; n=30) and non-native speakers (NNSs; n=30) of English. Six participants were excluded from our sample, because they had: (1) unstable internet connection, (2) low-quality microphones, and/or (3) poor (highly unintelligible) pronunciation. These problems negatively affect the quality of ASR output (e.g., Liakin et al., 2017). In addition, two native-speaker participants were excluded as they did not meet the main criterion for including in this group: only those born and raised in English-speaking countries and who used English as their first language were considered native speakers. The remaining 60 participants included undergraduate- and graduate-level students majoring in

different fields (e.g., Engineering, Education). See Appendix I for information about the participants' demographics and their previous experience with ASR.

Measures

Google's Voice Typing

Google's ASR engine, Google Voice Typing, was used in this study because it is free and readily accessible. Besides, Google's ASR is decently accurate, even to NNSs. Participants used Google Docs as the writing space to compose their texts.

Dictation Tasks

Participants were provided with two dictation tasks to practice transcribing with ASR. The tasks included two short, formal texts (i.e., leveled reading passages) adapted from the British Council (n.d.). The texts, which were the same for all participants, were at an English intermediate level (ranging from B1 to B2) according to the Common European Framework Reference for Languages (CEFR; https://www.coe.int). Participants were asked to transcribe the two pre-written texts by reading them aloud while the ASR system automatically transcribed them. The tasks helped participants become familiar with ASR-based writing, allowing for a more natural writing process that reflects the way we often think and speak, rather than the more structured process of typing.

Narrative Writing Tasks

Participants were provided with three narrative writing prompts adopted from Davis (2019); however, they were required to choose two of the prompts and use ASR to write their narrative texts. The tasks were designed as a precursor to the user perception survey, aiming to elicit firsthand experiences and perceptions from participants.

Narrative was chosen as a genre because discourse mode affects the writing performance. As Weigle (2010) asserted, "narrative … [is] often seen as cognitively easier and lend [itself] to less complex language than do exposition and argumentation" (p. 100).

Therefore, this study capitalized on using the narrative genre to direct participants' attention more towards text generation with ASR than complex linguistic components.

User Perception Survey

An online user perception survey adapted from the technology acceptance model 2 (TAM2; Venkatesh & Davis, 2000) was conducted to explore participants' (writers') perceptions of using ASR as a transcription mode. We employed the Google Forms survey platform for data collection. On average, participants spent approximately 20 minutes to complete the survey, which consisted of nine demographic questions and 29 statements with response options on a 5-point Likert scale: 1=strongly agree, 2=agree, 3=neutral, 4=disagree, 5=strongly disagree".

The overall aim of the survey is to measure participants' acceptability of using ASR as a transcription mode through seven constructs adapted from TAM2: Perceived ease of use (PEU), perceived usefulness (PU), intention to use (IU), attitude (AT), subjective norm (SN), output quality (OQ), and result demonstrability (RD). Each construct is represented in the form of some measurement statements that predict the acceptability of using the ASR system as a transcription mode, and that explain why the ASR system is adopted or declined to be adopted by the writer. See Appendix B for the full list of survey items.

Procedure

All the data collection procedures were carried out online via Zoom, a popular web-conferencing application. It is important to note that required measures were implemented to ensure the proper treatment of participants. As such, a comprehensive consent process was followed, wherein participants were fully informed about the study's aim, procedures, potential risks, and benefits. Participants had the chance to provide their informed consent before their involvement. In addition, all data were anonymized and securely stored throughout the research process. Finally, a compensation ($20) was provided for their time

and effort. The data collection included two main parts: *writing activities* and *user perception survey*. Writing activities consisted of two phases: *training* and *narrative writing*, both completed in Google Docs and via Google's ASR (Google Voice Typing). The session involving the writing activities was audio and video recorded to capture and analyze participant interactions with the technology. In our analysis, we focused on extracting insights from these recordings, particularly emphasizing each participant's writing processes, task performance, and any individual patterns observed.

Training Phase

At first, participants completed a free writing activity (i.e., writing about any topic for a short period of time – see Appendix C) to familiarize themselves with ASR. The free writing activity was employed as a warm-up task to assist participants in developing a writing mindset and overcoming any initial writer's block. Participants were asked to write freely (using ASR) without concerns over grammar or vocabulary choice because this activity was intended to be non-evaluative, thus promoting fluency and creativity in writing (Tynjälä et al., 2001).

The free writing activity was followed by a visual writing prompt, which was a picture of a landscape (see Appendix D). Participants were required to describe the picture and/or write about any ideas that the picture sparked in their mind. The free writing and visual writing activities served as a prelude to the more structured writing tasks (i.e., narrative writing tasks) and helped participants learn about how ASR responds to their speech input.

Finally, participants were provided with a dictation activity, including two short texts (see Appendix E). This activity helped participants practice writing longer texts with ASR and attune their writing behavior to punctuate and format their texts (e.g., adaptation to giving voice commands such as "Period", "New line").

For activities that did not require reading, participants were reminded that they were using speech as an input medium only; thus, they were strongly encouraged to avoid using features of spoken language in their texts (e.g., slang, colloquialism). Participants were also requested to consider several factors when inputting text to ASR: appropriate level of formality, punctuation, proper chunking (i.e., a seamless combination of words together as a phrase), and pronunciation intelligible to ASR. It is worth mentioning that in the context of ASR-based writing, free writing can be associated with dictation because both allow for uninterrupted process of producing text without undue concern about the mechanics of writing or the linguistic components of writing (e.g., grammatical accuracy).

Narrative Writing Phase

After the training phase, participants were asked to accomplish the narrative writing tasks (see Appendix F). This section included three narrative prompts; participants were required to choose two of them and spend 20 minutes on each to accomplish the tasks. Participants were advised to consider the three stages of the writing process: pre-writing (e.g., brainstorming), writing (e.g., focusing more on writing the first draft rather than on error correction), and post-writing (e.g., revising). After successfully completing the assigned tasks, participants were invited to complete a 20-minute survey based on the TAM2 model. The survey was conveniently conducted online using Google Forms, a widely recognized and trustworthy platform for conducting surveys (see Appendix G).

Result

To answer the research questions, this study adopted a quantitative approach using a user perception survey adapted from TAM2 (Venkatesh & Davis, 2000), using a 5-point Likert scale ranging from “strongly agree” (1) to “strongly disagree” (5).

Pilot Survey

The survey was piloted to ensure the internal consistency of the survey items (see Appendix H). Piloting was carried out with a sample size of 12 participants (NSs=6; NNSs=6). While literature recommends different sample sizes practical for pilot studies (e.g., between 10 and 30 participants - Isaac & Michael, 1995; or 10% of the total sample size - Lackey & Wingate, 1998), a sample size of 12 participants is considered reasonable for most pilot studies (Julious, 2005; Van Belle, 2002). The pilot survey was conducted using a different sample than that used for the main survey; however, the sample's demographics and education level were comparable to those of the main survey (see Appendix I). The purpose of the pilot survey was to identify any potential issues and, as such, refine the survey instrument before conducting the main survey, thus improving the validity and reliability of the main survey.

The survey used in this study consisted of seven constructs, including PEU, PU, IU, AT, SN, OQ, and RD. Therefore, to measure the internal consistency of every individual construct, seven Cronbach's alphas, one of the most common reliability tests, were run. It should be noted that Cronbach's alpha of 0.7 and above respectively shows a good and high level of internal consistency (DeVellis, 2003). As shown in Table 1, the values of Cronbach's alphas obtained in this study revealed a good and high level of internal consistency in all the seven constructs: PEU (α = .89); PU (α = .93); IU (α = .93); AT (α = .79); SN (α = .74); OQ (α = .83), and RD (α = .73). These coefficient values demonstrate that the survey has high reliability.

Table 3.1 illustrates the summary of average responses for each general TAM2 construct in the pilot survey. For instance, the mean value of PEU (M = 2.12) indicates that the average response to this item was mostly "Agree" (as "Agree" was coded as "2"). The r column shows corrected item-total correlation which is the correlation between a specific

item and the other items developed to measure a construct (See Appendix B for a detailed description of the reliability tests).

Table 3.1

Summary of Reliability Tests

TAM2 Construct	M	SD	*r*	Cronbach's α
Perceived Ease of Use (PEU)	2.12	1.10	.72	.89
Perceived Usefulness (PU)	2.80	1.27	.81	.93
Intention to Use (IU)	3.08	1.42	.84	.93
Attitude (AT)	2.41	1.07	.64	.79
Subjective Norm (SN)	3.52	1.20	.42	.74
Output Quality (OQ)	3.11	1.09	.69	.83
Result Demonstrability (RD)	2.19	1.13	.58	.89
Total	2.69	1.19	.69	.85

Writers' Perceptions of ASR as a Mode of Transcription (RQ1)

To answer the first research question regarding how writers perceive the use of ASR as a transcription mode, a one-sample t-test was conducted. It should be noted that participants were surveyed on seven perception-informed constructs, where anyone who achieved a score of 1.0 ("strongly agree") or 2.0 ("agree") was deemed to have positive perceptions of using ASR as a transcription mode. Higher scores indicated neutral (3.0) or negative perceptions (4.0 and 5.0).

A one-sample t-test was run to determine whether the perception score of participants was statistically significantly different from *neutral*, defined as a population perception score. Perception scores were normally distributed, as measured by Shapiro-Wilk's test ($p > .05$) and there were no outliers in the data, as assessed by inspection of a boxplot (see Figure 3.2). Tables 3.2 and 3.3 demonstrate that the participants' mean perception score ($M = 2.79$, $SD = .71$) was lower than the population average of 3.0, with a statistically significant mean difference of 0.20, 95% CI [0.01 to 0.38], $t(59) = -2.18$, $p = .033$. Therefore, the first null hypothesis that writers' perception average equals the neutral value of 3 was rejected.

However, as the Cohen's *d* effect size revealed, the mean difference was not practically significant, *d* = .28.

Figure 3.2

Perception Boxplot

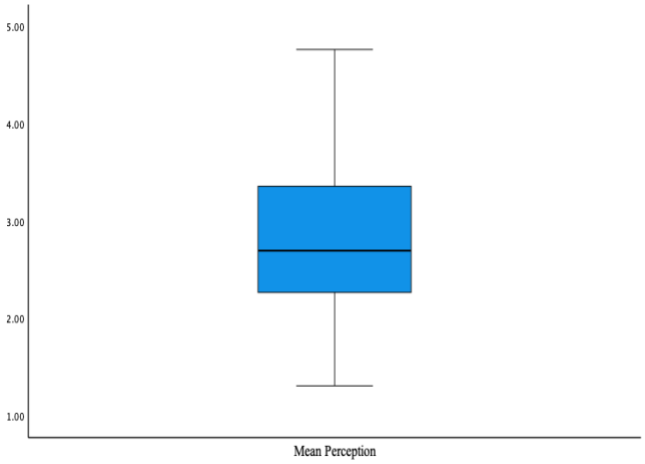


Table 3.2

One-Sample Statistics

	N	Mean	STD	Std. Error Mean
Writer Perception	60	2.79	.71	.092

Table 3.3

One-Sample T-Test Result

	Test Value = 3					
					95% CI of the Difference	
	T	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Writer Perception	-2.180	59	.033	-.20**	-.38**	-.01**

Tables 3.4 and 3.5 demonstrate the descriptive statistics and one-sample t-test results for every individual perception-informed construct investigated in this study. While these results show the participants' approval of some of the TAM2 constructs (i.e., PEU, AT, and RD), it can be maintained that, overall, the participants formed a neutral view towards the use of ASR as a transcription mode.

Table 3.4

Descriptive Statistics

Statistics		PEU	PU	IU	AT	SN	OQ	RD
N	Valid	60	60	60	60	60	60	60
	Missing	0	0	0	0	0	0	0
Mean		2.20	3.08	3.12	2.40	3.87	2.95	1.94
Median		2.16	3.00	3.12	2.33	4.00	3.00	1.75
Std. Deviation		.72	.90	1.05	.89	.88	.84	.97
Range		3	3.83	4	4	3	3.67	4
Minimum		1	1.17	1	1	2	1.33	1
Maximum		4	5	5	5	5	5	5

Table 3.5

One-Sample T-Test Results

	Test Value = 3					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
PEU	-6.399	59	.000	-.59722	-.7840	-.4105
PU	2.479	59	.016	.28889	.0557	.5221
IU	2.377	59	.021	.32500	.0514	.5986
AT	-3.453	59	.001	-.40000	-.6318	-.1682
SN	9.450	59	.000	1.07500	.8474	1.3026
OQ	1.434	59	.157	.15556	-.0615	.3726
RD	-6.791	59	.000	-.85833	-1.1112	-.6054

Comparing Native and Nonnative English Writers (RQ2)

To answer the second research question, which asked how native and nonnative English writers' perceptions compare in terms of using ASR as a transcription mode, a Mann-Whitney U test was conducted. As explained below (see also Table 3.6 and Figure 3.3), the results obtained indicate that there were no statistically significant differences between the perceptions of the two groups.

PEU. Distributions of PEU scores for NSs and NNSs were not similar, as assessed by visual inspection. PEU scores for NSs (mean rank = 37.43) were statistically significantly higher than for NNSs (mean rank = 23.57), $U = 242$, $z = -3.08$, $p = .002$.

PU. Distributions of PU scores for NSs and NNSs were not similar, as assessed by visual inspection. PU scores for NSs (mean rank = 32.62) and NNSs (mean rank = 28.38) were not statistically significantly different, $U = 386$, $z = -.942$, $p = .346$.

IU. Distributions of IU scores for NSs and NNSs were not similar, as assessed by visual inspection. IU scores for NSs (mean rank = 34.78) were marginally statistically significantly higher than for NNSs (mean rank = 26.22), $U = 321$, $z = -1.907$, $p = .056$.

AT. Distributions of AT scores for NSs and NNSs were similar, as assessed by visual inspection. AT scores were not statistically significantly different between NSs ($Mdn = 2$) and NNSs ($Mdn = 2.5$), $U = 390$, $z = -.886$, $p = .376$.

SN. Distributions of the SN scores for NSs and NNSs were not similar, as assessed by visual inspection. SN scores for NSs (mean rank = 34.10) and NNSs (mean rank = 26.90) were not statistically significantly different, $U = 342$, $z = -1.629$, $p = .103$.

OQ. Distributions of the OQ scores for NSs and NNSs were similar, as assessed by visual inspection. OQ scores for NSs ($Mdn = 2.5$) and NNSs ($Mdn = 2.5$) were not statistically significantly different, $U = 395$, $z = -.819$, $p = .413$.

RD. Distributions of the RD scores for NSs and NNSs were almost similar, as assessed by visual inspection. As Table 6 shows, the RD score was not statistically significantly different between NSs and NNSs ($Mdn = 1.5$), $U = 328$, $z = -1.844$, $p = .065$.

Overall, these findings indicate that NSs' and NNSs' views were similar in most of the TAM2 constructs, except PEU. Therefore, the second null hypothesis that English NSs will be more likely to perceive ASR positively than English NNSs was rejected.

Table 3.6

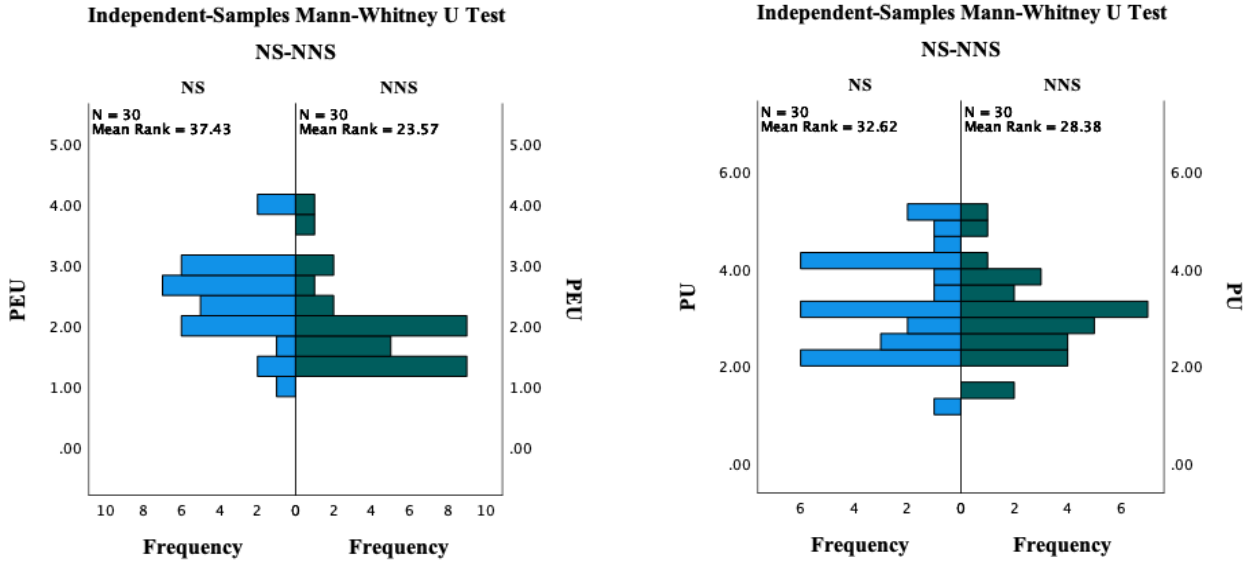
Independent-Samples Mann-Whitney U Test Results

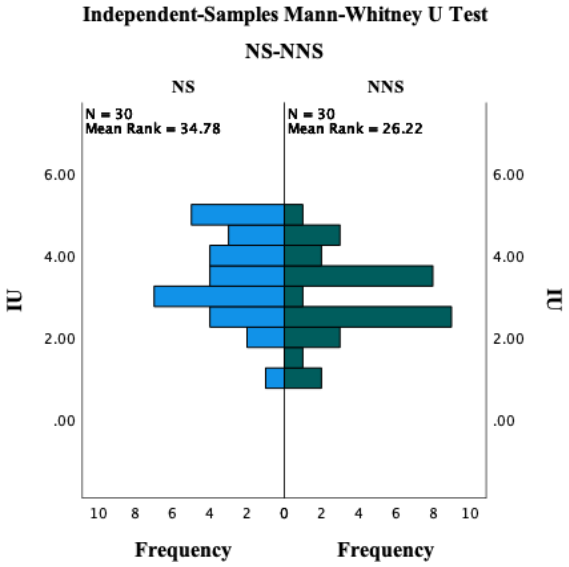
	PEU	PU	IU	AT	SN	OQ	RD
Total N	60	60	60	60	60	60	60
Mann-Whitney U	242.00	386.50	321.50	390.50	342.00	395.00	328.50
Wilcoxon W	707.00	851.50	786.50	855.50	807.00	860.00	793.50
Test Statistic	242.00	386.50	321.50	390.50	342.00	395.00	328.50
Standard Error	67.367	67.43	67.36	67.14	66.314	67.13	65.88
*STS	-3.088	-.942	-1.907	-.886	-1.629	-.819	-1.844
Asymptotic Sig.(2-sided test)	.002	.346	.056	.376	.103	.413	.065

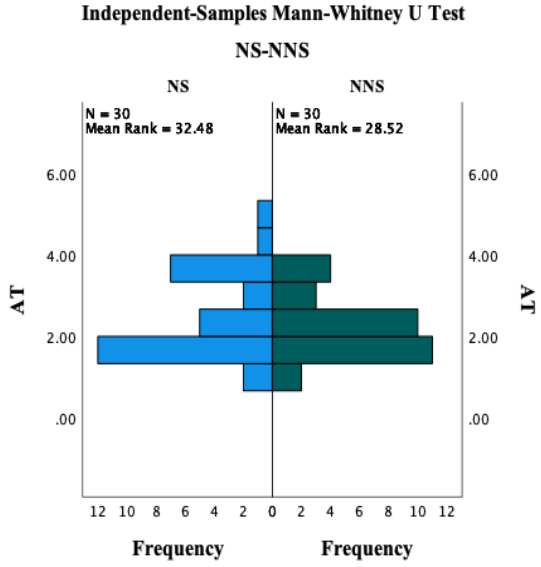
* Standardized Test Statistic

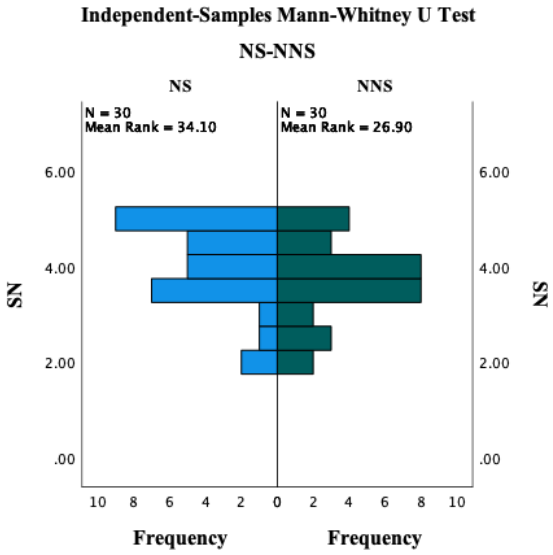
Figure 3.3

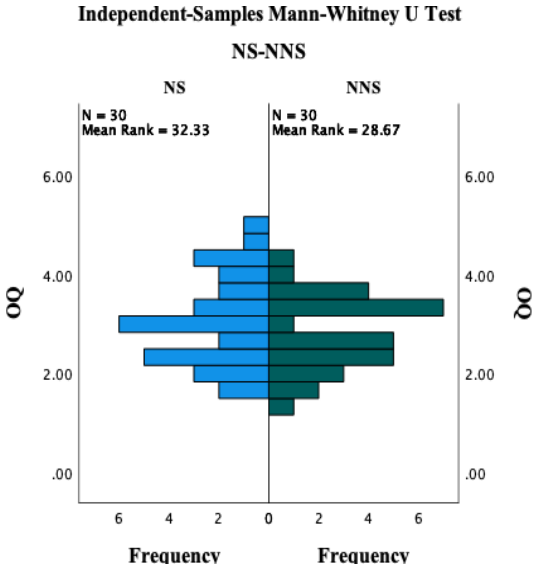
Distribution of Construct Scores

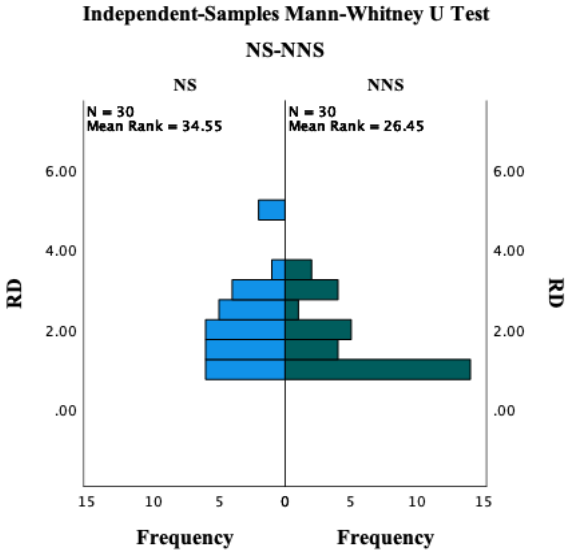



Independent-Samples Mann-Whitney U Test
NS-NNS
NS
NNS
N = 30
Mean Rank = 34.78
N = 30
Mean Rank = 26.22
IU
Frequency
Frequency


Independent-Samples Mann-Whitney U Test
NS-NNS
NS
NNS
N = 30
Mean Rank = 32.48
N = 30
Mean Rank = 28.52
AT
Frequency
Frequency


Independent-Samples Mann-Whitney U Test
NS-NNS
NS
NNS
N = 30
Mean Rank = 34.10
N = 30
Mean Rank = 26.90
SN
Frequency
Frequency


Independent-Samples Mann-Whitney U Test
NS-NNS
NS
NNS
N = 30
Mean Rank = 32.33
N = 30
Mean Rank = 28.67
OQ
Frequency
Frequency


Independent-Samples Mann-Whitney U Test
NS-NNS
NS
NNS
N = 30
Mean Rank = 34.55
N = 30
Mean Rank = 26.45
RD
Frequency
Frequency

Discussion

This study investigated writers' perceptions of using the ASR technology as a transcription mode for composing narrative texts. Sixty participants (30 NSs and 30 NNSs) were recruited and took part in an ASR-based experiment (in which they were asked to compose texts using their voices), and then in a user perception survey. Guided by TAM2 (Venkatesh & Davis, 2000a), the survey found evidence for how writers perceived the use of ASR as a transcription mode, and how English NS's versus NNS's perceptions of using this technology as a transcription mode compare.

Concerning the first research question, which asked how writers perceive the use of ASR as a transcription mode, a one-sample t-test was conducted to determine whether perception level in participants was different from *neutral* (i.e., the hypothesized mean value in the population). The results revealed that participants' perception average was lower than the population average, meaning that participants mostly *agreed* with or were *neutral* towards accepting ASR as a transcription mode. Around 71.5% and 22.5% of participants respectively expressed agreement with and neutrality towards ASR being *easy to use*. This can be explained by the intuitive interface of ASR (van Doremalen et al., 2016). Contrary to previous research that highlights the *usefulness* of ASR for improving pronunciation (Chen, 2011; Inceoglu et al., 2023), only 27.5% and around 34% of participants respectively agreed with and were neutral towards the *usefulness* of ASR for writing. The low agreement with PU can be explained by participants' lack of (a) automaticity with voice commands (Wetzel, 1996), e.g., for punctuation; and (b) ASR engine's accuracy in recognizing voice commands as actual commands, not as text, e.g., typing "period" instead of the punctuation mark "." (Young & Mihailidis, 2010).

Additionally, 31.7% of participants indicated that they would *intend to use* ASR for writing; however, 34% were neutral as to whether they would continue using ASR over

keyboard typing. One explanation for this slightly low intention rate is that Google's ASR engine sometimes outputs misrecognized words (e.g., target word "fluent" misrecognized as "affluent"), which can be easily overlooked in the ASR's output because these two words are orthographically similar. Therefore, the writer must be overly attentive to such errors, which can slow down the process of writing (Leijten & Van Waes, 2005).

Interestingly, participants had a positive (60%) and neutral (approximately 30%) *attitude* about using ASR for writing. This can be explained by the ASR's input medium, which is speech-based. As a natural and convenient means of communication, speech is often developed faster and earlier than writing (van Doremalen et al., 2016). Further, speech promotes fluency in transcription as it reduces or removes transcription issues (MacArthur & Cavalier, 2004), thus reducing the writer's cognitive load for writing tasks (Quinlan, 2004). Consequently, the writer using ASR can efficiently produce text, which is essential to writing development (McCutchen et al., 1994).

Participants agreed (10%) or were neutral (20.85%) about the influence of *subjective norm* on accepting ASR as a transcription mode. As defined by Fishbein and Ajzen (1975), recall that this construct refers to the user behavior on using or refusing a tool, determined by the influence of important people such as a teacher or school board. According to our findings, it can be claimed that subjective norm is not a strong determinant of accepting ASR as a transcription mode.

With reference to *output quality*, 38.4% and 28.3% of the participants respectively expressed agreement with and neutrality towards the quality of their produced texts. One explanation for this slightly low satisfaction rate is that most participants were either unfamiliar with ASR or with writing long texts using their voices. Another possible explanation may be due to the duration and format of the treatment (e.g., without spaced

practice), which did not give participants enough opportunities to improve their ASR-based writing skills.

Regarding *result demonstrability*, participants agreed with (73%) and were neutral towards (20.8%) the discernability and communicability of the interrelationship between ASR use and positive outcomes. This positive finding can be justified by ASR's main affordance: the use of one's voice as an input medium, which possibly enhanced the writers' ability to produce more text (Leijten & Van Waes, 2006; Quinlan, 2004).

Let us now focus on the second research question, which asked how NSs and NNSs English writers' perceptions compare in terms of using ASR as a transcription mode. The results of the Mann-Whitney U test revealed that both NSs and NNSs had almost similar perceptions towards using ASR as a transcription mode, except for PEU. These comparable perceptions can be explained by ASR's output quality. As the results showed, new ASR systems, particularly the one utilized in this study (Google), are not significantly affected by L2-accented speech (McCrocklin & Edalatishams, 2020), so NNSs writers could equally capitalize on the ASR recognition accuracy and produce text as could NSs.

However, NSs obtained a higher PEU score than NNSs, indicating that they perceived ASR easier to use than their non-native counterparts. One possible explanation may be due to the fact that NSs can input text more fluently than NNSs (e.g., without long pauses or hesitations). As reported by some participants, long pauses often affect Google's ASR responsiveness. To troubleshoot this problem, the user must enable and disable the system upon each unresponsiveness so that it can work properly – a cumbersome approach to composing texts, which likely affected NNS's perceptions of ASRs vis-à-vis ease of use.

Limitations

The present study suffers from a number of limitations. One major limitation is that, as indicated in the participants' background questionnaire, most had never previously used

ASR for writing long texts (n=52), and among those who had used the technology (n=8; e.g., to write text messages), their writing behavior was not attuned to using voice for producing larger texts. Observing the participants' interactions with ASR, one of the most common takeaways was that they were not fully prepared to format their texts (e.g., to add punctuations, start a new paragraph), which hindered the flow of text generation for most writers and, most likely, affected their perceptions of ASR use for composing texts. Besides, a single intervention session may not have provided sufficient time for participants to develop and refine the skill of inputting text via ASR. Another limitation is that the participants interacted with ASR while being recorded and observed by one of the researchers, so it is possible that they experienced writing anxiety, which could have affected their overall perceptions. Furthermore, this study was conducted using a one-sample design, which limits the generalizability of the results to other populations or contexts.

Conclusion

This study has shown that ASR systems such as Google's can be viewed as a writing tool for both NSs and NNSs English writers because the system optimally understands and accurately transcribes both native and non-native L2 speech. Our findings add to the existing body of knowledge by providing insights into the (a) writer perception of ASR use as a transcription mode, an area that has received very little research, and (b) differences in perception between NSs and NNSs English writers' use of ASR for writing, which was neglected in earlier research. Additionally, the results of this study can inform the development and enhancement of ASR systems for writing purposes. Moreover, the study broadens our understanding of how ASR affordances can improve writing instruction for language learners. For instance, as discussed before, ASR use as a transcription mode fosters a process approach to writing, which has been proven to be effective in developing writing skills (Hayes, 1996). Furthermore, teachers can use ASR as a new (or alternative)

transcription mode to motivate their students (with or without learning disabilities) to produce drafts of their texts more efficiently, consequently promoting their language production using a resource that they already possess—their own speech. Students can then be encouraged to read and review their texts for further revisions and edits, thus promoting writing as a process, not as the final product.

Future Directions

We suggest that future research could conduct a similar study with a larger and more diverse sample to improve the generalizability of the findings. In addition, future research should consider designing studies with longitudinal interventions (e.g., as part of an English writing curriculum) so that participants can automatize the necessary skills to write effectively and efficiently and, as such, we can have a more reliable picture of writers' perception, preferably after the novelty effect (Clark, 1983) has worn off.

Chapter Four: Exploring Second Language Writers' Usage Behavior of Automatic Speech Recognition: A Focus on Lexical Diversity in Narrative Texts

While speaking is a natural mode of communication that doesn't require external tools such as keyboards or pen and paper, writing, on the other hand, is an activity that relies on the use of external tools. Thus far, different input methods (e.g., typewriters, ballpoint pens, keyboards) have been developed for writing purposes. One of the technological innovations in the development of writing (the "voice evolution"; Enge, 2020) is automatic speech recognition (ASR), an accessible and ubiquitous technology that converts users' speech to text, which can be found in many current devices (e.g., laptops, smartphones) and in word-processing and communication software (e.g., Google Docs, Microsoft Office Suite, text messaging applications). As such, ASR has the potential to revolutionize the way individuals can effortlessly transform their spoken language into written form, expanding the possibilities of written communication.

Despite the ongoing increase in the use of voice as an input tool in mobile human-computer interaction (Enge, 2020), as well as the extensive use of ASR in the second language (L2) pronunciation research, ASR-assisted writing has not received the attention it deserves. Although research shows the efficacy of ASR technology in improving L2 pronunciation (Liakin et al., 2015) and describes the strategies that L2 learners adopt when practicing pronunciation with the technology (Dillon & Wells, 2021; Van Lieshout & Cardoso, 2022), the exploration of ASR-assisted writing remains relatively limited, with only a small number of studies delving into this area. Moreover, the majority of these studies have focused on individuals with learning difficulties (Leijten & van Waes, 2005; MacArthur, 1999). Importantly, very few studies have investigated (1) how writers interact with ASR to produce text, regarding their behaviors; and (2) whether there is a relationship between the writers' choice of behavior and the lexical diversity of their final text.

To address these research gaps, this study employed an exploratory sequential design to identify the behaviors that participants employ when writing narrative texts via ASR (e.g., exclusive use of ASR, combining keyboarding with ASR) and to provide a preliminary examination of the relationship between the behaviors used and the diversity of lexical items in their texts. An exploratory sequential design, a mixed methods research approach, involves initially collecting and analyzing qualitative data. The results then inform the development of subsequent quantitative research, allowing for a comprehensive and iterative understanding of research questions (Johnson & Onwuegbuzie, 2004). To contextualize and provide motivation for the study, the component skills involved in the writing process and the significance of learner behavior for learning (and writing) will be briefly reviewed below.

Background

Importance and Challenges of Writing

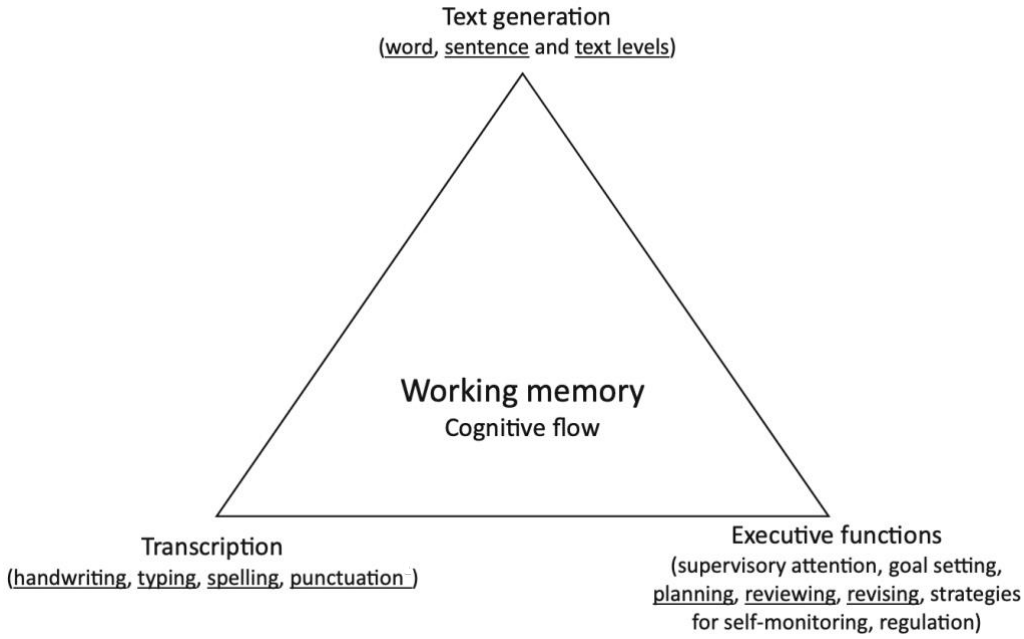
Effective writing skills play an important role in fostering literacy and developing communicative competence, and these skills are frequently assessed in high-stakes exams as a means to ensure one's academic and professional capabilities (Weigle, 2010). However, research has demonstrated that learners' writing skill is deteriorating (Healey & Gardner, 2021). As Silva (1993) explains, L2 writing tends to be problematic because it is generally more restricted, more challenging, and less successful than L1 writing; this is owing to the idea that L2 writers "plan less, revise for content less, and write less fluently and accurately than first language writers" (Weigle, 2010).

Challenges in L2 writing can be explained through the lenses of the developmental model of writing process (Berninger & Winn, 2006). Based on this model, three interconnected component skills are engaged in working memory throughout the production of text: *text generation*, *transcription skills*, and *executive functions* (see Figure 4.1). Text generation is described as the writer's capability to utilize vocabulary and grammar to

translate ideas into linguistic representations (i.e., language) at the word, sentence, and discourse level (Berninger et al., 1992). Transcription skills are defined as "the translating of … language representations in working memory into written symbols on the printed page" (Berninger et al., 1996). Finally, executive functions are viewed as the writer's ability to self-regulate the cognitive and metacognitive processes involved in writing (Berninger & Amtmann, 2003). Executive functions comprise "conscious attention, planning, reviewing [or monitoring], revising, [and] strategies for self-regulation" (Berninger & Amtmann, 2003).

Figure 4.1

Developmental Model of Writing (Adopted from (Berninger et al., 2009)



These component skills can impose constraints on the composing process, thus prompting the writer to develop strategies to facilitate the writing task. It should be noted that the first focus of this study is to probe learner behaviors (e.g., a subcomponent of learning strategies) associated with *executive functions* (research question 1). So, let us first describe the term 'learner behavior'.

Learner Behavior

Learner behavior refers to the various actions, strategies, and processes that language learners use when acquiring a second language (Griffiths, 2007). Some key aspects of learner behavior include learning *strategies*, which will be the focus of this study. Griffiths (2007) describes strategies as specific actions that learners deliberately take to simplify task completion (e.g., rehearsing, memorizing vocabulary). Oxford (1999) defines the concept as "specific actions taken by the learner to make learning easier, faster, more enjoyable, more self-directed, more effective, and more transferable to new situations" (p. 8). In this research, learner behavior serves as an overarching term encompassing a spectrum of strategies and processes employed by language learners in their acquisition of a second language. This inclusive term encapsulates the diverse approaches and cognitive mechanisms learners utilize throughout their language learning (Chaudron, 1988). Within this broad framework, strategy use becomes a focal point, representing the intentional and goal-driven approaches that learners employ during various stages of language acquisition. As Cohen (2014) emphasizes, this is particularly evident in the early stages of performing a novel task or using a new technological tool. When used frequently, these strategies identified earlier become familiar and might be utilized automatically (Chamot, 2005).

Learning strategies can be divided into four major interrelated types: Cognitive, metacognitive, affective, and social. According to Stewner-Manzanares et al. (1984), *cognitive* strategies refer to "the direct manipulation or transformation of learning materials in order to enhance learning and retention" (p. 9). In the context of ASR writing, examples of cognitive strategies include vocabulary selection (i.e., writers' word and phrase selection during dictation to ensure accurate recognition by the ASR) and syntax monitoring (i.e., structuring spoken sentences to align with proper written grammar and punctuation rules). *Metacognitive* strategies include learners' reflection on their thought processes. ASR writers

apply metacognitive strategies to regulate their writing process by planning (i.e., setting clear goals for the writing process), self-monitoring (i.e., reflecting on the quality of the transcriptions), and problem-solving (i.e., identifying and addressing issues related to transcription errors) (Stewner-Manzanares et al., 1984). *Affective* strategies are "emotional or attitude-related" such as identifying and using a method to lower anxiety during the writing process (Oxford, 1986). In ASR writing, affective strategies encompass attributes such as motivation (i.e., maintaining a positive attitude and staying motivated to utilize ASR for writing), self-efficacy (i.e., believing in one's ability to use ASR proficiently), and emotional regulation (i.e., managing frustration or impatience in instances where the technology may not work perfectly) (Oxford, 1999). Finally, *social* strategies generally refer to learning with and from others. Examples pertaining to ASR-based writing include collaborating with others, such as peers or editors, who can assist with reviewing and improving ASR-generated text, as well as seeking feedback from teachers to hone the quality of transcriptions (Ma & Oxford, 2014; Oxford, 1986).

Writing Behaviors and Individual Differences

Guided by the abovementioned definitions and categorizations, writing behaviors (WBs) refer to the actions that the writer takes to overcome the constraints laid by component skills (i.e., text generation, transcriptions, and executive functions) to facilitate the writing process. In traditional writing activities, a range of behaviors can be employed to enhance the writing process. These behaviors include strategies such as brainstorming ideas, outlining concepts via mind-mapping, drafting and revising the text, utilizing resources such as thesaurus or dictionaries, and conducting content revisions. However, the factors influencing language writers in their selection of these strategies are not always apparent beyond the instructions provided by teachers and textbooks. What is clear, though, is that the choices made by L2 writers are likely influenced by their individual differences (for similar claims,

see Dörnyei, 2005; Pawlak & Kruk, 2022). According to Dörnyei (2005) and Pawlak & Kruk (2022), individual differences refer to the unique characteristics, abilities, preferences, and needs of language learners, which can influence their interactions with technology and their overall language learning outcomes. Dörnyei (2005) categorizes these differences into three main categories: Language Aptitude (cognitive abilities such as working memory and analytical skills), Motivation (the degree of desire, effort, and commitment towards obtaining a learning goal), and Personality Traits (characteristics that shape an individual's behavior and preferences).

This study also utilizes theories pertaining to cognition and multimedia learning to conceptualize the WBs that L2 learners might use. One of them is *cognitive flexibility*, which refers to an individual's capacity to transition between various cognitive processes or task. This theory focuses on how people change their thinking and behavior in response to new conditions, tasks, or expectations (Spiro et al., 1991). L2 learners' use of WBs can also be conceptualized using Sweller's (1988) *cognitive load theory*. This theory is concerned with how the cognitive demands of a writing task affect learners' ability to successfully produce text. It implies that employing WBs such as planning and chunking information that distribute cognitive load can assist L2 writers to enhance their writing performance. Finally, this study uses the concept of "linearity" to justify the adoption of specific WBs (Papert, 1980). The term linearity refers to a sequential progression from one tool to another, such as transitioning from exclusive use of ASR to exclusive use of keyboarding while composing texts.

In the context of ASR-assisted L2 writing, exploration of WBs is essential for two main reasons: First, it provides us with insight into the cognitive, metacognitive, affective, and social processes that ASR writers are usually involved with during the speak-to-write process. Second, it enables us to impart new WBs (i.e., introducing novel strategies) to less

competent or novice ASR writers to improve their writing skills (Ardasheva & Tretter, 2013; Grenfell & Harris, 1999). Research indicates that higher strategy use leads to improved L2 proficiency (Ardasheva & Tretter, 2013; Lan & Oxford, 2003), enhanced academic achievement in subject areas such as language arts (Chamot, 2007), and improved cognitive and behavioral performance (Magogwe & Oliver, 2007). Another focus of this study is to examine whether the behaviors participants use while writing with ASR affect the lexical diversity of their final text (research question 2).

Lexical Diversity

Lexis (or vocabulary) is a key component of language, both spoken and written (Nation & Newton, 2008). Lexical diversity is described as a measure of the number of distinct lexical items used in a text. It is regarded as an informative determinant of learners' overall language skill (G. Yu, 2010) and an important indicator of their writing quality (Laufer & Nation, 1995). The significance of lexical diversity is also demonstrated in the writing scoring scale of many international language proficiency texts (e.g., IELTS, TOEFL). For example, IELTS writing band descriptors consist of four criteria by which written responses are assessed, and one of them is 'lexical resource' (i.e., 'the range of vocabulary used') that constitutes 25% of the total writing score (*IELTS Guide*, 2022). Moreover, in terms of measuring text features, lexical diversity has been found to be one of the most essential elements in automated writing assessment systems (McNamara et al., 2014).

As Laufer and Nation (1995) indicate, lexical diversity is an important factor in determining the vocabulary employed in a written text and, as a result, the holistic text quality assessment, particularly if the text is written by an L2 learner with a limited vocabulary in comparison to native speakers. For example, in a study by Linnarud (1986), the lexical diversity of texts written by L2 learners (Swedish learners of English) was lower than that of texts written by English native speakers, which has an adverse effect on the overall

quality grading of learners' written texts. Due to the importance of lexical diversity in L2 writing, this study employs four indices to assess the lexical diversity of participants' texts (see Method for further details). The indices are variations of type-token ratio (TTR), which is one of the most widely employed methods for measuring lexical diversity. TTR is calculated by dividing the number of unique words (types) in a text by the total number of words (tokens) used. The resulting indices were used to examine whether the behaviors participants use in an ASR-assisted writing environment affect the lexical diversity of their final texts.

ASR-assisted Writing

The term ASR-assisted writing refers to the act of producing text with speech recognition tools (e.g., Google Voice Typing, Microsoft Dictate). While there has been a substantial body of research conducted on ASR-based interventions in providing feedback and creating practice opportunities for L2 learners to improve their pronunciation skills (Inceoglu et al., 2023; Liakin et al., 2015), ASR-assisted writing remains relatively underexplored, despite several benefits it can provide to support writing. This gap in research is further amplified in the context of adult L2 learners, particularly among those with no learning and/or writing difficulties. Below, we will describe some of the benefits that ASR-assisted writing provides.

Inclusive Writing Environment

New ASR systems afford an *inclusive* writing environment because their sensitivity to accented L2 speech is relatively low (McCrocklin & Edalatishams, 2020) and, as such, L2 learners with different pronunciation proficiency levels can conveniently use the system to input text without experiencing too many recognition errors. Further, people with learning (and writing) disabilities can easily use the system for writing purposes. Research indicates that ASR can contribute to the writing development of people with learning disabilities

(Caute & Woolf, 2016). Interestingly from a pedagogical standpoint, ASR does not promote one specific writing behavior, as users can maintain their planning, reviewing, or pausing behavior, developed by other input tools (e.g., keyboards) in ASR-assisted writing (Leijten & van Waes, 2005).

Speech-based Input Tool

ASR technology serves as a *speech-based input* tool, offering L2 learners a means to enhance their text production efficiency. By harnessing the natural and rapid nature of speech (Selouani et al., 2008), the technology promotes fluency in the writing process. Compared to typing, speech-based input proves to be a faster medium for language production (Quinlan, 2004). In addition, using speech as an input method reduces reliance on keyboards and mitigates challenges related to typing, such as poor spelling and keyboarding difficulties (MacArthur, 1999).

The Developing Text

Finally, ASR facilitates the creation of *the developing text*, which refers to the text displayed on the screen immediately after speech input. This feature fosters process-based writing by allowing the writer to continuously engage in reading, reviewing, planning (the remaining text), and revising the content throughout the composing process (Leijten & Van Waes, 2005). Research has underscored the importance of rereading, reviewing and revising the developing text in writing development (Flower & Hayes, 1980; Hayes, 1996).

The Current Study

Despite the multiple affordances that ASR offers to writers, its potential for L2 writing as well as its effects on the lexical diversity of texts have not been widely explored. Given that there is scant evidence available on what happens when L2 writers produce text using ASR (particularly regarding the behaviors employed), and how these behaviors affect the linguistic components of the text (e.g., lexical diversity), it is important to explore users'

behaviors when producing text via ASR, and the association between lexical diversity and L2 writers' choice of behavior(s).

Cardoso (2022) proposed a chronological framework for describing research in CALL consisting of four stages: (1) Development: the development of a new technological tool (not the focus of this study because ASR technology has already been developed); (2) Exploration: refers to the examination of pedagogical affordances and potential of the proposed technology to address a number of concerns (e.g., user behavior with ASR) by assessing how they meet practical and/or theoretical criteria for the effective use of the technology; (3) Assessing suitability: entails research on learners' perceptions (level of acceptance) of using the chosen technology; and (4) Assessing pedagogical effectiveness: emphasizes the technology's ability to foster learning a target skill. Since the present study is exploratory and causal in nature, the researcher will focus on stage 2 (exploration: research question #1) and stage 4 (assessing pedagogical effectiveness: research question #2).

Numerous studies have shown the pedagogical benefits of ASR in improving L2 pronunciation (Evers & Chen, 2021; Liakin et al., 2015; Mroz, 2018). However, only a small number of research has explored the impact of using ASR as a writing tool – i.e., as a tool for inputting text (exceptions include: Macarthur, 2000; Quinlan, 2004; Reece & Cummings, 1996). Following recommendations by Chapelle (2010), who call for studies exploring learners' behaviors when using technology with the aim of establishing a relationship between the strategy used and outcomes (in this study: the lexical diversity of ASR-written texts), two research questions are proposed:

RQ_1: How do writers interact with ASR to produce text, in terms of behaviors employed?

RQ_2: Is there an association between the ASR writers' choice of behaviors and lexical diversity of their final text?

Method

Participants

This study recruited 30 adult participants (mean age: 28 years old) through random sampling. Participants were English L2 speakers, and they were undergraduate- and graduate-level students in different academic domains (e.g., Computer Science, Education), with a relatively equal distribution by gender: male (46%) and female (54%). All participants were proficient in speaking and writing in their first languages. Approximately 95% of participants received their first writing instruction in their first language.

Participants consisted of people with different first languages (e.g., Chinese, Farsi, French, German, Hindi, Portuguese, Vietnamese), and about 67% of them had lived in English-speaking countries for four years or more. Regarding their experience using ASR, nine participants had previously used tools such as Siri and Alexa for improving their speaking and pronunciation, and for asking questions using search engines. Only one participant had previously used ASR for texting and writing emails (see Appendix I for further details on participants' demographics and their previous experience with ASR). To ensure the anonymity of the participants, a numerical coding system was employed. Each participant was assigned a unique numeric code (e.g., P3) instead of using their actual identities.

Measures

Google Voice Typing

The Google Voice Typing, an ASR system developed by Google, was employed in this study because it is cost-free and easily accessible. Additionally, Google Voice Typing is fairly accurate, even for non-native speakers (McCrocklin et al., 2019). Each participant was provided with a blank page on Google Docs, an online word processor, to create their texts.

Writing Tasks

Three narrative writing prompts, adapted from (Davis, 2019), were given to the participants; however, they were asked to select two of the prompts and utilize ASR to compose their narrative texts.

The motive for choosing narratives was determined by Weigle's (2010) view that this genre or "discourse mode" correlates highly with writing performance. According to Weigle (2010), "narrative … [is] often seen as cognitively easier and lend [itself] to less complex language than do exposition and argumentation" (p. 100). Likewise, (Crowhurst, 1980) indicates that narratives often evoke shorter T-units (text units that include one independent clause as well as any dependent clauses linked to it) than do argumentative texts. As a result, the present study opted for examining narratives due to its relative simplicity in terms of production. This choice allowed our participants to concentrate on utilizing ASR for text generation and on devising strategies to tackle challenges associated with ASR-assisted writing.

Procedure and Data Analysis

All data collection procedures were conducted online using Zoom, a widely used web-conferencing application. It is essential to highlight that rigorous measures were in place to protect participants' rights. A thorough consent process was administered, ensuring participants were fully briefed on the study's objectives, procedures, potential risks, and benefits. Participants had the opportunity to give informed consent before engaging in the study. Furthermore, all collected data were anonymized and securely stored throughout the research process. To acknowledge their time and effort, participants received compensation ($20).

An exploratory sequential design was adopted to answer the research questions. As such, following Geertz's (2008) method of *thick description*, this study examines participants'

behaviors while using ASR to produce text, in terms of behaviors used. Thick description is a way of providing in-depth explanations of participants' perspectives underpinning their observable behaviors within a particular context (Geertz, 2008). Guided by a holistic approach for the analysis of users' behaviors and experience (Gil Urrutia et al., 2017), this study analyzed data gained through the video recordings and screen captures of participants' interactions with ASR, as well as the researcher's fieldnotes.

Video Recordings and Screen Captures

The participants' interactions with ASR were monitored and analyzed using video recordings and screen captures to explore different behaviors that they used while composing text. The identified behaviors were coded through *inductive* codes. As Johnson and Christensen (2019) describe, "inductive codes … are generated by the researcher by directly examining the data during the coding process" (p. 548). The coding method enabled the researchers to discover the ways in which the participants interacted with ASR concerning behaviors used and to determine whether there is an association between the behaviors employed and the lexical diversity of ASR-written texts.

Indices of Lexical Diversity

In this study, the lexical diversity of the participants' texts (n = 30) was analyzed. The texts were imported to the Tool for the Automatic Analysis of Lexical Diversity (TAALED; Kyle et al., 2021), an analysis tool developed to compute a wide range of lexical diversity indices. TAALED, one of the most widely used software in the field of writing assessment (Zenker & Kyle, 2021), is open-source and free to use, featuring an intuitive interface that is compatible with both Mac and Windows operating systems. For this study, as mentioned earlier, four indices of lexical diversity that have been found to be relatively resistant to text length were selected, including Root type-token ratio (Root TTR; Guiraud, 1960), moving-average TTR (MATTR; Covington & McFall, 2010), the measure of textual lexical diversity

(MTLD; McCarthy & Jarvis, 2010), and MTLD moving-average-wrap (MTLD-MA-W; McCarthy & Jarvis, 2010). Each is operationally described below.

Root TTR. This index is calculated by dividing the number of *types* (i.e., unique words in a text) by the square root of the number of *tokens* (i.e., total number of words in a text). For instance, if the verbs *take*, *took*, *taken*, and *taking* are present in a text, they would be considered as one type, but four different tokens. Texts with higher Root TTR scores are more lexically diverse (Guiraud, 1960).

MATTR. The moving average type-token ratio (MATTR) is obtained by averaging TTRs across many chunks of text that are overlapping and evenly sized. For instance, for a 20-word chunk, TTR values are determined for tokens (words) 1–20, 2–21, 3–22, etc., until every token in the text appears in at least one chunk, and the final MATTR score is produced by averaging all the TTR values. Texts with higher MATTR scores are more lexically diverse (Kyle et al., 2021).

MTLD (Original). The measure of textual diversity (MTLD) is an index that indicates the average number of words required to achieve a certain TTR value (the default value = .720). Each word in a text is sequentially evaluated for its TTR during the MTLD computation. To ensure the words at the end of the text have achieved the default value of .720, MTLD runs forwards and backwards through a text and computes TTR on word strings of increasing lengths (called 'factors'). As such, two MTLD scores are computed: One for the forward computation and one for the backward. The final MTLD score is calculated by averaging the two scores (McCarthy & Jarvis, 2010).

MTLD-MA-Wrap. This index is a revised version of MTLD that utilizes a moving average approach to calculate lexical diversity. MTLD-MA-Wrap prevents partial factors (i.e., the final words that do not reach the TTR value of .720) by moving back to the

beginning of the text rather than proceeding through the text forwards and backwards (Zenker & Kyle, 2021).

Results

Strategies Used by L2 Writers When Interacting with ASR

Adopting an exploratory approach**,** this study analyzed the collected video recordings and screen captures to answer the first research question on how writers interact with ASR to produce their text regarding behaviors used. As will be reported next, each participant used a variety of behaviors, which are broadly categorized into four types: (1) *ASR exclusively (ASR-∀*, with the universal quantifier symbol "∀" to differentiate the strategy from the technology), (2) *ASR in tandem with keyboarding (ASR=KB)*, (3) *ASR followed by keyboarding (ASR>KB)*, and (4) *ASR followed by both keyboarding and ASR (ASR>ASR=KB)*.

ASR Exclusively (ASR-∀)

Two participants (out of 30; /30 henceforth) produced text exclusively using ASR (ASR-∀ henceforth) for all stages of writing, including planning, drafting, revising, and editing. For example, regarding planning, the participants used a number of voice commands in place of graphic organizers (e.g., a hierarchical organizer, a bubble map) to show the connection between ideas. The commands given consisted of: "period" to separate ideas from one another; and "new line", "insert number", or "create bulleted list" to arrange and classify ideas throughout the planning phase. In terms of editing, the participants used the command "select [word or phrase]" to select the error word(s) then spoke the appropriate term to replace the selected one. For instance, during the intervention, P14 used ASR to produce the following sentence: "However I had the privilege to attend some French classes", which lacked a comma after the transition word 'however'. P14 asked ASR to "select [however]" then spoke to it, saying 'however,' in which she repeated the transition word followed by a

comma to edit and fix the punctuation issue. These examples demonstrate how P14 carried out different stages of writing using only ASR.

ASR in tandem with Keyboarding (ASR=KB)

Thirteen participants (/30) used ASR in tandem with the keyboard when producing text (ASR=KB). The simultaneous use of ASR and the keyboard was performed to revise and edit text at every level of writing (e.g., drafting, revising). Participants in this category preferred to edit their sentences (and paragraphs) before proceeding to the next one and completing their texts. For instance, using ASR, P3 wrote "he was 24 years old he was an accountant" and immediately turned to the keyboard to change it to "He was 24 years old when I met him. He was an accountant". In this example, P3 used ASR to create a sentence that missed the first-word capitalization, proper use of punctuation marks, and the conjunction "and". P3's output also lacked sentence complexity. She addressed these issues by immediately using the keyboard to complement the ASR output, for instance, by capitalizing the letter "h" in 'he', adding an adverbial clause 'when I met him' to increase the complexity of the sentence, and inserting a period to mark the end of the sentence.

ASR followed by Keyboarding (ASR>KB)

Five participants (/30) utilized ASR only to complete the initial drafts of their texts; all further revisions and editing were done using the keyboard exclusively (ASR>KB). Following this behavior, the participants composed their first drafts relatively fast because they did not go back to fix problems. After completing their first drafts, these five participants exclusively used their keyboards (with ASR turned off) to correct errors and edit their texts. For example, using ASR to create the first draft, P8 wrote the following sentence "sometimes you *are having like* a good conversation sometimes we *saw* and it was on and off but suddenly I *realize* that's not a *white* relationship" (incorrect words are in italics to facilitate processing). Then, using his keyboard, P8 revised the sentence as "sometimes we

had a good conversation and sometimes we found out that it was intermittent but suddenly, I realized that's not a right relationship". As the revised text shows, P8 minimized spoken colloquialism by removing the word 'like', increased the sentence formality by changing 'saw' to 'found out' and 'on and off' to 'intermittent', and corrected verb tenses by altering 'are having' to 'had' and adding the past marker -ed to 'realize'. P8 also added the missing punctuation and corrected the ASR recognition error by changing 'white' to 'right'.

ASR followed by both ASR and Keyboarding (ASR>ASR=KB)

A final writing behavior identified among ten participants involved the exclusive use of ASR for generating the initial draft, followed by a combination of keyboarding and ASR for inputting and making corrections to the text (ASR>ASR=KB). For instance, with ASR, P30 wrote: "*And* the beginning of 2020, I started learning a very different language […] *Interesting Lee*, French was not my mother tongue". After finishing the initial draft, P30 asked the ASR to "select [And the beginning of]" and then said "backspace" to delete the selected words, thus substituting the deleted phrase with the preposition 'in'. P30 then used his keyboard to replace "*Interesting Lee*" with the intended "Interestingly".

Table 4.1 summarizes the strategies adopted by the participants and their distribution in the data.

Table 4.1

ASR-based Behaviors and Their Distribution

Behaviors	n	%
ASR-∀	2	7
ASR=KB	13	40
ASR>KB	5	17
ASR>ASR=KB	10	33

Associations between Behaviors and Lexical Diversity

To answer the second research question regarding whether there is an association between the ASR writers' choice of behaviors and lexical diversity of their final written text,

a two-step procedure was used. First, the lexical diversity of participants' texts was calculated using TAALED, software developed by Kyle et al. (2021) to measure the level of lexical diversity in a text. TAALED provides measurements of the text's diversity in terms of its lexicon, capturing the range and variety of words used.

Table 4.2 shows the descriptive statistics for the four indices used to determine the lexical diversity of the texts analyzed. The t-scores along with the mean values of all four indices computed are displayed in Table 4.3. A t-score is a standardized score that indicates how distant an observed value is from the mean value. It ranges from 20 to 80, with a standard deviation of 10 and an average of 50. According to the mean t-scores obtained, a total of 21 narrative texts had lexical diversity equal to or greater than average. The $n \geq 50$ for each index consists of: Root TTR (n=19); MATTR (n=16); MTLD Original (n=21); MTLD-MA-Wrap (n=20).

Table 4.2

Descriptive Statistics for the Indices Calculated

Index	N	Min	Max	M	SD
Root TTR	30	35.73	67.25	51.26	9.65
MATTR	30	34.01	67.09	50.87	9.40
MTLD	30	27.58	70.25	53.19	11.59
MTLD-MA-Wrap	30	31.29	70.79	52.69	11.14

Table 4.3

T-scores of Indices Measured

Participants	Root TTR	MATTR	MTLD	MTLD-MA-W	M
P1	58.37	55.28	62.52	60.24	**59.10**
P2	36.22	38.74	40.23	39.58	38.69
P3	58.95	57.91	63.89	64.54	**61.32**
P4	36.15	34.01	27.58	34.02	32.94
P5	65.28	63.59	68.87	66.47	**66.05**
P6	51.89	52.91	57.65	58.33	**55.20**
P7	53.56	50.55	51.98	54.85	**52.74**
P8	55.48	59.73	51.95	50.67	**54.46**
P9	37.22	42.46	39.97	39.81	39.87
P10	46.20	60.00	51.68	50.40	**52.07**
P11	48.69	46.83	54.21	55.88	**51.40**
P12	54.91	48.19	56.82	52.84	**53.19**
P13	54.91	58.56	63.74	61.06	**59.57**
P14	50.83	47.83	60.01	57.40	**54.02**
P15	50.08	47.63	52.84	49.82	**50.09**
P16	49.84	50.19	55.91	57.18	**53.28**
P17	38.45	39.74	41.86	42.65	40.68
P18	53.99	52.91	54.89	54.32	**54.03**
P19	58.45	59.55	66.75	63.38	**62.03**
P20	67.25	64.91	70.25	69.23	**67.91**
P21	35.73	36.37	33.23	31.34	34.17
P22	44.29	41.1	48.02	46.40	44.95
P23	62.34	62.37	63.60	60.92	**62.31**
P24	36.72	41.1	33.89	31.29	35.75
P25	59.37	52.91	51.55	52.58	**54.10**
P26	37.07	38.74	36.68	34.31	36.70
P27	57.10	48.19	61.20	62.57	**57.27**
P28	53.85	43.46	44.38	47.99	47.42
P29	61.42	63.19	69.65	70.79	**66.26**
P30	63.33	67.09	59.93	59.91	**62.57**

Bolded numbers indicate mean values equal to or above average.

Next, Eta (η) and Eta-squared (η^2) tests were conducted to determine the association between the writing behaviors used and the lexical diversity of the written narratives. Eta is a measure indicating the strength and direction of the association between the independent and dependent variables, whereas Eta-squared is an effect size measure indicating how much of the variance in the dependent variable can be explained by the independent variable.

As Table 4.4 shows, five levels of lexical diversity were identified in the collected narratives, ranging from 'below average' (<50) to 'above average' (>50). Our analysis

reveals that while the two narrative texts produced using ASR-∀ obtained below-average lexical diversity, the use of ASR=KB resulted in texts with lexical diversity that was slightly above (*n*=7) and above average (*n*=4). Texts using ASR>KB were rated as having below-average (*n*=3), slightly below-average (*n*=1), and slightly above-average (*n*=1) lexical diversity. Lastly, regarding the use of ASR>ASR=KB, most of the texts showed lexical diversity levels that were slightly above average (*n*=4) and above average (*n*=4).

Table 4.4

Writing Behaviors and Associations with Lexical Diversity

		Levels of Lexical Diversity					
		Below	Slightly below	Average	Slightly above	Above	Total
WS	ASR-∀	2	0	0	0	0	2
	ASR=KB	1	1	0	7	4	13
	ASR>KB	3	1	0	1	0	5
	ASR>ASR=KB	0	1	1	4	4	10
Total		6	3	1	12	8	30

According to the Eta and Eta-squared results (see Table 5), there is a moderately strong association between the use of the ASR-∀ and the low degree of lexical diversity (ASR-∀; η = .4). Given the obtained Eta-squared value, the lexical diversity of the collected texts was significantly impacted by applying ASR-∀, $F(29) = 6.80$, $p < .05$, $\eta^2 = .19$. The large effect size found ($\eta^2 = .19$) reveals a meaningful effect between the variables (for further details on effect sizes, see Cohen, 1992). In addition, the use of ASR combined with keyboarding is somewhat strongly associated with the high level of lexical diversity in the texts (ASR=KB; η = .2). The obtained Eta-squared value denoted a negligible effect of ASR=KB use on the lexical diversity of the texts, with a medium effect size detected, $F(29) = 1.77$, $p > .05$, $\eta^2 = .05$.

Our results also show that the adoption of ASR followed by keyboarding is moderately strongly associated with a low degree of lexical diversity (ASR>KB; η = .4).

According to the Eta-squared result, the lexical diversity of the texts was significantly affected by employing ASR>KB, with a large effect size obtained, $F(29) = 8.99$, $p < .05$, $\eta^2 = .24$. Lastly, writing with ASR followed by both keyboarding and ASR is moderately linked to a high degree of lexical diversity (ASR>ASR=KB; $\eta = .3$). As demonstrated by the Eta-squared result, the adoption of this behavior does not significantly impact lexical diversity, with a large effect size reported, $F(29) = 3.68$, $p > .05$, $\eta^2 = .11$.

The effect size reported indicates that there is no meaningful relationship between the behavior used and the texts' lexical diversity. It is important to acknowledge that although our findings, particularly those pertaining to ASR>KB and ASR>ASR=KB, failed to attain statistical significance, the presence of large effect sizes implies that there may be meaningful or practical significance if a larger sample size is employed. The findings of the Eta and Eta-squared tests are summarized in Table 4.5.

Table 4.5

Summary of the Measures of Associations and their Effect Sizes

WS	*F*	*p*	η	η^2
ASR-∀	6.80	.014	.44	.19
ASR=KB	1.77	.194	.24	.05
ASR>KB	8.99	.006	.49	.24
ASR>ASR=KB	3.68	.065	.34	.11

In summary, our findings reveal that using ASR-∀ corresponds to a moderate negative effect on lexical diversity with a significant effect, while ASR=KB is associated with a moderate positive relationship but with no statistically significant effect. The adoption of ASR>KB is strongly correlated with a low level of lexical diversity, with a significant effect. Finally, the use of ASR>ASR=KB exhibits a moderately medium positive relationship with high lexical diversity, but no statistically significant impact is detected.

Discussion

This study investigated how participants interact with ASR regarding the behaviors that they employ to produce narrative texts using this technology (RQ1), and whether the behaviors employed affect the lexical diversity of the participants' written narratives (RQ2). Thirty English NNSs participated in an intervention in which they were asked to create texts utilizing ASR and other writing tools at their disposal (e.g., keyboards, pen and paper). Our findings demonstrate that (1) participants employed four writing behaviors to produce text utilizing ASR, and (2) certain behaviors (i.e., ASR=KB, ASR>ASR=KB) are more likely to be associated with the production of higher quality texts, as they yielded greater lexical diversity.

Behaviors Used in ASR-based Writing

An exploratory analysis was carried out concerning the first research question, which asked what behaviors L2 writers use when interacting with ASR to produce texts. Using video capture recordings and field notes, four distinct behaviors were observed: (1) ASR exclusively (ASR-∀); (2) ASR in tandem with keyboarding (ASR=KB); (3) ASR followed by keyboarding (ASR>KB); and (4) ASR followed by both keyboarding and ASR (ASR>ASR=KB).

The presence of a diverse range of behaviors suggests that writers have their own writing (and possibly learning) preferences, which stem from individual differences (IDs), including different cognitive abilities and learning styles (e.g., visual, auditory). Dörnyei (2005) defines IDs as characteristics or attributes that may distinguish individuals from each other. As such, the two participants who exclusively employed ASR-∀ for composing and editing their texts might have heightened auditory and oral abilities in comparison with their peers; this led them to rely on ASR throughout all phases of the writing process, including

planning, drafting, revising, and editing (for a discussion of these types of learners in SLA, see Pawlak & Kruk, 2022).

The prevalence of ASR=KB for writing among the thirteen participants could potentially be indicative of their cognitive flexibility (Spiro et al., 1991). This flexibility allowed the participants to tailor personal characteristics to best accommodate the task at hand: for instance, using ASR when verbal expression is more efficient, and keyboarding when precision is required (e.g., when editing).

The five participants who utilized ASR>KB can be classified as "linear" writers. In line with the concept of linearity (Papert, 1980), the participants initially relied on producing text using ASR due to its suitability for their communication style and for overcoming the challenge of generating text (Weigle, 2010). However, as their writing needs became more complex and demanding, they switched to keyboarding to obtain more precise control over the writing process (e.g., via revising and editing).

Lastly, the second most commonly used behavior was ASR>ASR=KB, adopted by ten participants. These participants possibly relied on their working memory capabilities (i.e., the capacity to retain and manipulate information within working memory), an ID attribute that can affect how writers manage cognitive load. According to Sweller's (1988) Cognitive Load Theory, working memory has a finite capacity, and when it is overloaded, learning can be hindered. In the case of creating initial drafts in L2 writing, switching between the ASR and keyboard may cause cognitive overload and hinder the writing process. Thus, we hypothesize that the participants' use of this strategy reflects their relatively low working memory capacity, which was alleviated by use of ASR-∀ in the initial stages of the writing process.

ASR-associated Behaviors and Lexical Diversity

Regarding the second research question, Eta (η) and Eta-squared (η^2) tests were conducted to investigate the connection between the writing behaviors employed and the lexical diversity of the resulting written narratives. Our findings indicate that the utilization of the ASR combined with keyboarding, namely ASR=KB and ASR>ASR=KB, resulted in a high level of lexical diversity. This finding supports the modality effect theory brought forth by Leahy and Sweller (2011), which postulates that using multiple input modalities, such as visual and auditory channels, might assist to evenly disperse the cognitive load, resulting in more effective writing outcomes. With the use of these behaviors, participants could develop and refine ideas and avoid the cycle of constant text monitoring and recurring corrections in texts created with ASR-∀, thereby resulting in the production of more lexically diverse texts. Likewise, it can be claimed that certain writers may adopt this behavior because they are given more flexibility to use different input modalities to create text.

Another theory that could explain the abovementioned results is the Cognitive Theory of Multimedia Learning proposed by Mayer (2005). According to this theory, learners can process information in two separate channels: a visual/pictorial channel and an auditory/verbal channel. The concurrent use of both methods, as observed in ASR=KB and ASR>ASR=KB, could potentially activate both channels simultaneously, resulting in an increased level of lexical diversity. In the case of ASR-∀ and ASR>KB, on the other hand, writers may have only engaged one of these channels at a time, which could explain the low degree of lexical diversity observed in the texts produced by individuals who utilized these two behaviors.

Overall, these results consistently demonstrate that the use of ASR and keyboarding in tandem (ASR=KB) may lead to an increase in the lexical diversity of the text. This suggests that the integration of ASR technology alongside traditional keyboarding methods

effectively promotes a broader lexical range and variety in written expression. These findings underscore the potential of leveraging ASR in conjunction with keyboarding as a valuable tool for fostering greater lexical diversity in written communication.

Limitations and Future Directions

Given the exploratory nature of this research and the inherent limitations associated with a study that explores uncharted territory, it is crucial to acknowledge some limitations. One of the most important limitations is its small sample size, which may restrict the generalizability of our findings to a broader population. Future research could use a larger sample size to enhance the external validity of the findings. Another limitation regards the short duration of this study since the experiment was not longitudinally designed. Future research may examine how the writing behaviors described in this study ultimately impact lexical diversity over an extensive period of ASR-based writing. Concerning the second research question, since there was no control group in the study, it is difficult to attribute the observed effects solely to the use of the behaviors used by the participants. Future studies should include a control group to isolate the impact of the selected behaviors more clearly on lexical diversity. Finally, future research could also consider the relationship between the writing behaviors and other measures of texts (e.g., sentence complexity) written by ASR.

Conclusion

The findings of this study suggest that L2 writers adopt a variety of behaviors for inputting text when interacting with ASR, highlighting the presence of individual differences in their approach. Additionally, the employment of various input modalities (e.g., ASR, keyboarding) can impact the lexical diversity of narrative texts. According to our findings, combining ASR with keyboarding (ASR=KB or ASR>ASR=KB) enhanced lexical diversity, while the exclusive use of ASR (ASR-∀) or ASR followed by keyboarding (ASR>KB) did not.

Overall, learning behaviors that can increase lexical diversity (e.g., those that emphasize the combined use of ASR and keyboarding) can contribute to improvements in the development and instruction of ASR-assisted L2 writing. For instance, teachers can leverage ASR as a potentially powerful tool to motivate their L2 learners in the realm of writing. By combining ASR with keyboarding during writing instruction, teachers can offer learners a more efficient way to draft their texts. This not only enhances the writing process itself, but also fosters a sense of active engagement and productivity among the learners.

Chapter Five: Conclusion & General Discussion

Automatic speech recognition systems such as Google's can be utilized as a writing tool for English writers due to the systems' ability to recognize and transcribe both native and non-native L2 speech. As such, ASR can streamline the writing process by removing (or minimizing) the necessity for manual typing. Writers can employ ASR systems to dictate their ideas, receive immediate text feedback, and then revise and edit the text as needed. The use of ASR can be beneficial in a variety of ways. Teachers can effectively utilize ASR as an alternative writing tool to foster motivation among their students, regardless of whether they have learning disabilities or not. By leveraging ASR technology, students are empowered to create text drafts more efficiently by tapping into a resource they already possess – their own speech.

ASR-facilitated writing can also enhance productivity and attentiveness towards the process of text generation, as the technology can offer the added advantage of enabling students to seamlessly integrate writing into their daily activities. The technology's mobility facilitates user freedom, allowing students to produce text without being tied to a specific location. This flexibility empowers students to engage in writing anytime and anywhere, breaking the barriers of traditional writing constraints. As a result, writing assisted by ASR becomes more flexible and adaptable to students' lifestyles and schedules, enhancing their overall writing experience and productivity.

The three manuscripts included in this dissertation explores the use of ASR for writing and provides some insight into user perceptions and the pedagogical potential of using ASR for writing. Each manuscript addressed one aspect of the overarching research questions: (a) What are the pedagogical benefits and limitations of ASR use for L2 pedagogy and for writing purposes? (Manuscript A; Chapter 2); (b) What are learners' perceptions of

ASR use as a writing tool? (Manuscript B; Chapter 3); and (c) What strategies do learners use when engaged in ASR-assisted writing? (Manuscript C; Chapter 4).

In response to these questions, this chapter will provide an overview of each manuscript, a summary of major findings, and a discussion of the pedagogical implications of employing ASR for writing. It will also discuss the limitations of the dissertation, directions for future studies, and concluding remarks.

Overview, Summary, and Pedagogical Implications of the Manuscripts

Guided by Cardoso's (2022) chronology for CALL research, this dissertation investigated the pedagogical potential of ASR use for writing. Since ASR systems have already been developed, this dissertation bypassed the first stage ('Development') and proceeded to the subsequent stages of CALL development – stages 2 through 4 – with each chapter delving into one of these stages. Firstly, it delved into stage 2, which involved exploring the pedagogical potentials and limitations of ASR in the context of L2 pedagogy, with a particular emphasis on its impact on writing skills. It then proceeded to stage 3 by evaluating the suitability of ASR as a writing tool based on users' perceptions and experiences. Lastly, stage 4 encompassed an examination of ASR's potential to promote the adoption of effective writing strategies and how these strategies relate to the lexical diversity observed in ASR-written texts.

Manuscript A - Chapter 2: A Systematic Literature Review of Automatic Speech Recognition in L2 Learning: A Case for L2 Writing

The aim of this study was to systematically review the benefits and limitations of using ASR in L2 education and as a writing tool. ASR research has been mainly concerned with the development of oral skills, particularly speaking and pronunciation (Cardoso, 2022; Cox & Davies, 2012) and, to a lesser extent, with using ASR as a writing tool for people with writing (or learning) disabilities (Quinlan, 2006). The existing body of research on using

ASR as a writing tool for individuals without learning disabilities is quite limited, with only a small number of studies conducted approximately two decades ago (Leijten & Van Waes, 2005; Reece & Cummings, 1996). Furthermore, there is a notable lack of studies exploring the use of ASR as a writing tool specifically for adult L2 English writers. Consequently, it can be stated that this systematic review is the first effort in investigating the advantages and limitations of ASR in the context of L2 writing development. By filling this research gap, this study shed some light on the unexploited potential of ASR technology in enhancing the writing skills of L2 English writers.

This study justified the use of ASR in L2 education utilizing the software evaluation framework (SEF; Burston, 2003; Hubbard, 1988) and insights from Second Language Acquisition (SLA; e.g., Output hypothesis – Swain, 1985). Following Hubbard (2006) and the CALL literature (e.g., Chapelle, 2001; Egbert & Shahrokni, 2018), this study systematically reviewed 79 empirical studies published between 1990 and 2021 to examine the pedagogical potential and limitations of ASR use for L2 learning, particularly for writing purposes. This study was guided by the following question: What does the empirical literature on ASR use in L2 education reveal about the potential benefits and limitations of the technology in terms of (a) L2 education in general; and (b) L2 writing specifically?

To answer the research question, the thematic analysis conducted found repeated themes concerning the potential benefits and limitations of using ASR in L2 education as well as L2 English writing. The findings revealed that the pedagogical use of ASR is commonly attributed to benefits such as improving pronunciation, providing opportunities for L2 output practice and explicit teaching, the creation of an inclusive learning environment, and the tool's accessibility, particularly regarding cost-effectiveness. The findings, however, suggest a number of limitations such as the low recognition accuracy of the system, especially for accented non-native speakers, and ASR's inability to provide explicit

corrective feedback. Moreover, despite the human-like interaction that ASR affords (Cardoso, 2022), it was found that ASR-based interaction lacks certain essential elements of effective oral communication; notably, the absence of nonverbal cues was highlighted as a significant limitation (for similar claims, see also Mroz, 2018).

In the context of writing, the results revealed that the most frequently mentioned benefit of ASR is its capacity to serve as a speech-based input tool, promoting a process-oriented approach to writing. ASR also supports an inclusive writing environment by enabling writers of different abilities to engage in the writing process. However, our analysis indicates that the use of ASR for writing purposes is not without limitations. One notable limitation identified is the tool's spelling accuracy, which may result in users overlooking recognition errors in the written text. Given this limitation, there is a clear need for an improved error correction mechanism in current ASR applications. In addition, the findings suggest that text formatting can be challenging when using ASR tools for writing. As ASR technology primarily converts speech into text, the process of formatting the written text, such as applying headings, bullet points, or other structural elements, may not be as seamless or intuitive compared to traditional text input methods. This limitation can affect the overall visual presentation and organization of the written text. Finally, the systematic review conducted in this study serves as a foundation for the next chapters. While Chapter 3 (Manuscript B) delves into understanding L2 writers' perception of the using ASR as a writing tool, Chapter 4 (Manuscript C) examines the strategies used by writers when using ASR to compose their narrative texts and the correlations between the strategies used and the quality of their output.

Manuscript B - Chapter 3: Users' Perceptions of ASR as a Writing Tool: An Analysis using the Technology Acceptance Model

Informed by the technology acceptance model 2 (Venkatesh & Davis, 2000), this study aimed to explore users' perceptions of utilizing ASR for writing to demonstrate whether they are willing to accept it as a transcription mode. The study also compared native and nonnative English writers' perceptions of using ASR as a writing tool. To attain these aims, the following questions were addressed.

1. How do writers perceive the use of ASR as a transcription mode?
2. How do native and nonnative English writers' perceptions compare in terms of using ASR as a transcription mode?

Using a quantitative one-sample design, this study recruited 60 adult participants, stratified among two groups: Native speakers (NSs; n=30) and non-native speakers (NNSs; n=30) of English. Participants completed a training phase, including a free writing activity and a visual writing prompt to become familiar with ASR-based writing. After the training, they were asked to complete three target narrative writing tasks. After having used ASR to compose texts, participants were invited to fill out the TAM2-based survey.

Regarding the first research question, a one-sample t-test was conducted to determine the perception level of participants on a number of TAM2-informed items. The results revealed that participants mostly agreed with or were neutral towards accepting ASR as a transcription mode. The majority of participants (71.5%) expressed that ASR was easy to use, whereas a small percentage of participants (27.5%) believed that ASR was useful for writing. Consequently, only 31.7% of participants indicated that they intended to use (or continue to use) ASR for writing. Overall, a high percentage of participants (60%) had a positive attitude toward using ASR for writing.

Considering subjective norm (one of the variables that predicts users' acceptance of a new technology, according to TAM2), the results indicated that a relatively small percentage of participants, specifically 10%, agreed that this variable had an influence on the acceptance of ASR as a transcription mode. This suggests that subjective norms, which refer to individuals' perceptions of pressures to use a particular technology (e.g., from peers, from their teacher), may not significantly impact the acceptance of ASR as a preferred mode of transcription. In terms of output quality, approximately 38.4% of the participants believed that the texts they produced using ASR were of good quality, indicating that a significant portion of the participants perceived positive outcomes in terms of the quality of their written texts when writing with ASR. Finally, a substantial majority of the participants (73%) agreed with ASR's capacity to demonstrate a discernable relationship between the use of ASR and positive results. This finding suggests that the majority of participants recognized the potential of ASR technology to yield positive outcomes.

Concerning the second research question, a Mann-Whitney U test was conducted to assess the participants' perception of ASR in terms of willingness to use it and ease-of-use, and how these variables compared among native vs. non-native English writers. The results showed that both NSs and NNSs had similar perceptions when it came to employing ASR as a transcription mode. This suggests that the use of ASR technology, specifically the system utilized in the study (Google), is not significantly influenced by the presence of a non-native accent. As a result, both NS and NNS writers were able to effectively utilize ASR to create text. However, the perceived ease-of-use score of NSs was higher than that of NNSs, suggesting that NSs found ASR to be more user-friendly compared to their non-native counterparts. This indicates that NSs may have perceived ASR technology as more intuitive and easier to use in the context of transcription.

Based on these findings, it can be concluded that ASR systems are generally perceived positively as a writing tool for both NSs and NNSs of English. However, while the study's findings provide valuable insights into the perceptions and experiences of users, it is necessary to conduct further research to fully understand and evaluate the effectiveness of using ASR to produce texts. Thus, following Cardoso's (2022) chronology for investigating CALL tools, Chapter 4 (Manuscript C) investigated the strategies employed by users when utilizing ASR to generate texts, and whether there exists a correlation between the type of strategy adopted and the quality of the produced text, specifically in terms of lexical diversity.

Manuscript C - Chapter 4: Exploring Second Language Writers' Usage Behavior of Automatic Speech Recognition: A Focus on Lexical Diversity in Narrative Texts

The goal of this last manuscript was to identify the behaviors that writers use when writing with ASR and to provide a preliminary analysis of how the lexical diversity of written texts and the behaviors employed by users relate to one another. Guided by stage 2 (exploration) and stage 3 (assessing pedagogical effectiveness) of Cardoso's (2022) chronology for CALL research, this study employed a mixed-methods approach to investigate users' behavior in the context of creating narrative texts using ASR. The research involved analyzing the behaviors utilized by writers during the text creation process and examining the association between their choice of behavior(s) and the lexical diversity observed in their texts. This study addressed the following two questions:

1. How do writers interact with ASR to produce text, in terms of behaviors employed?
2. Is there an association between the ASR writers' choice of behaviors and lexical diversity of their final text?

For this study, a total of 30 L2 English-speaking participants were recruited. These participants consisted of undergraduate and graduate-level students who were pursuing their

studies in various academic domains, such as Computer Science and Education, among others. Each participant received a blank page on Google Docs (an online word processor) to write their texts using Google Voice Typing. The participants were given three narrative writing prompts to write about. Using video recordings and screen capture software, the participants' interactions with ASR were observed and recorded, and later analyzed to explore their text-composing behaviors. Indictive coding was used to code the identified behaviors.

The lexical diversity of the participants' texts was examined using the Tool for the Automatic Analysis of Lexical Diversity (TAALED; Kyle et al., 2021), an analysis tool developed to measure a wide range of lexical diversity indices. For this research, four lexical diversity indices were selected, including Root type-token ratio, moving-average TTR, the measure of textual lexical diversity, and MTLD moving-average-wrap.

Regarding the first research question, the study employed an exploratory analysis using field notes and screen recordings of participants' screen activities. According to this analysis, four different behaviors were found, in which participants used: (1) ASR exclusively (ASR-∀), (2) ASR in tandem with keyboarding (ASR=KB), (3) ASR followed by keyboarding (ASR>KB), and (4) ASR followed by both keyboarding and ASR (ASR>ASR=KB).

Concerning the second research question, Eta (η) and Eta-squared (η2) tests were conducted to investigate the association between the writing behaviors used and the lexical diversity of the written narratives. The findings obtained indicate that integrating ASR with keyboarding (ASR=KB or ASR>ASR=KB) may result in a higher lexical diversity in the text. This shows that the combination of ASR technology with keyboarding effectively fosters written expression with a wider lexical range and variation. These results highlight the

possibility of using ASR in tandem with keyboarding as a useful technique for developing more lexical diversity in texts.

This study makes an important contribution to CALL by highlighting the potential benefits of utilizing ASR technology in L2 writing instruction and the importance of incorporating specific writing behaviors alongside ASR. These results also contribute to our understanding of L2 writers' experiences regarding the impact of different writing behaviors, both individually and in combination, on ASR-based writing. L2 teachers can utilize ASR technology as a powerful tool to motivate and engage their students in the writing process, promoting more efficient drafting of their written work by leveraging their own speech. By encouraging students to utilize ASR for generating initial drafts, teachers can foster increased written language output and facilitate a more dynamic writing experience. Furthermore, after using ASR to produce their initial drafts, students can be guided to transition to other input media, such as a keyboard, to engage in the editing and revision stages of the writing process. This approach encourages a process-oriented writing approach, where students focus on refining their written work through thoughtful editing and revision.

Highlights

The three studies presented above provide valuable insights into the pedagogical potential of ASR in the writing process. These studies emphasize how ASR technology can enhance the accuracy and efficiency of transcription, allowing writers to harness their existing speech skills to create written text, reducing (or even eliminating) the need for manual typing.

The first manuscript examined the pedagogical benefits and limitations of ASR in L2 learning and writing. The findings shed light on the positive aspects of ASR as well as its limitations, providing a comprehensive understanding of its role in language education. One of the main pedagogical benefits highlighted in this paper is the promotion of autonomous

learning, as ASR enables learners to independently practice their speaking and writing skills by providing instant feedback and transcription assistance. This fosters self-directed learning as students can actively engage with the technology and receive immediate guidance on their language production. However, this study also identified a number of limitations of ASR use, particularly its low recognition accuracy. ASR systems may encounter challenges in accurately transcribing non-native speech, especially when it includes accents or pronunciation variations. This limitation can impact learners' experience and effectiveness in using ASR for language learning purposes.

The second manuscript revealed important insights into the perception of ASR systems as writing tools, specifically focusing on Google Voice Typing. The findings underscored that ASR systems are generally regarded as user-friendly and hold the potential for usefulness in the writing process. Additionally, this study shed light on ASR's ability to instill a positive attitude among users. An important contribution of this study was its exploration of the perceptions of both native and non-native English speakers regarding the use of ASR for writing. Findings revealed that there was no significant difference in the perceptions of these two groups, suggesting that both native and non-native English writers view ASR as a viable tool for writing. Overall, these findings suggest that the perception of ASR as a writing tool is not significantly influenced by the writer's native or non-native English speaker status, further emphasizing the technology's universality and potential for enhancing writing experiences across diverse language backgrounds.

Finally, the third manuscript explored the impact of incorporating ASR in the writing process, specifically focusing on the behaviors adopted by L2 writers and their potential correlation with the quality of their textual output. Findings revealed that the utilization of ASR in writing facilitated the adoption of a number of behaviors for text production, including the use of ASR exclusively, ASR followed by subsequent editing and revising

through keyboarding, as well as a combination of ASR and keyboarding. The results also revealed possible correlations between the behaviors adopted by L2 writers and the quality of their textual output (e.g., behaviors that involve the combination of ASR with keyboard such as ASR=KB, ASR>ASR=KB are more likely to be associated with texts of higher quality, with greater lexical diversity), suggesting that the choice of behavior may influence the overall linguistic richness of the produced text. By examining these writing behaviors, this study provided valuable insights into how writers navigate the use of ASR to optimize their writing process.

Limitations and Future Directions

While the three studies included in this dissertation have provided valuable insights into the pedagogical use of ASR in L2 writing, it is important to acknowledge that they are primarily exploratory in nature. Conducting pioneering research often involves venturing into new territory and exploring uncharted areas. As a result, these studies serve as a starting point for further investigation and open avenues for future research in the field.

Being exploratory in nature, these studies have inherent limitations. The first limitation relates to the small sample sizes and representativeness across the studies covered (manuscripts B and C), which limits the generalizability of the findings. Additionally, the participants may have come from specific academic domains and age groups (adults, university educate), which may not fully reflect the diversity of L2 learners. Expanding the sample size and including participants from a wider range of backgrounds would enhance the external validity of the findings. The second limitation of this research is its exclusive reliance on one ASR system, Google Voice Typing, which may have its own limitations in terms of accuracy, compatibility, or user interface. Considering different ASR technologies (e.g., Microsoft Dictate, Apple Siri) could provide a more diverse assessment of the technology's capabilities and limitations. The third limitation to consider is the potential

influence of the novelty effect (Clark, 1983), which refers to the initial excitement and interest that participants may experience when using a new technology or tool. In the case of ASR as a writing tool, the novelty effect may have contributed to the predominantly positive perceptions reported in manuscript B. It is important to acknowledge that the novelty effect may diminish over time, as participants become more accustomed to using ASR as a writing tool. Extended use of ASR in writing tasks may lead to a reduction in the initial excitement and novelty, potentially impacting participants' perceptions and experiences. Therefore, it is crucial to conduct further research to investigate the long-term effects and sustainability of using ASR as a writing tool. This will help to understand whether the positive perceptions and benefits observed during the initial stages of adoption remain consistent and durable over extended periods of use. Finally, the research was constrained by the short duration of the intervention and writing activities, which lasted approximately two hours. As it is widely recognized, writing, especially when using a less commonly employed input tool (ASR), can be a laborious process. The limited time allocated for interaction and practice with ASR may have impacted the participants' ability to fully develop and refine their ASR-based transcription skills. Expanding the duration of the intervention would have allowed for increased interaction and practice with the technology, enabling participants to gain a deeper understanding of its functionalities and potential benefits. This, in turn, could have led to a more comprehensive exploration of the pedagogical potential of ASR in L2 writing.

Concluding Remarks

This dissertation examined the use of a popular automatic speech recognition system, Google Voice Typing (Google's speech recognition engine), among native and non-native adult English writers to determine its efficacy as a writing tool for producing narrative texts. The research was conducted in three stages: first, it explored the pedagogical benefits and limitations of ASR based on CALL/SLA literature (Manuscript A); second, it investigated

users' perceptions of utilizing ASR for writing purposes (Manuscript B); and third, it examined users' behaviors when compositing texts with ASR and possible correlations between the behavior(s) adopted and text quality (Manuscript C). The findings demonstrate that ASR systems as those found in Google Voice Typing have the potential to enhance transcription efficiency (and possibly accuracy), enabling writers to generate text using a tool they all possess: their voice.

Despite the need for further exploration in various aspects of ASR use for both L1 and L2 writing, this dissertation offers evidence that supports the pedagogical effectiveness of ASR as a tool for writing tasks and sheds light on the potential benefits of the technology in writing instruction. The findings of this research pave the way for future investigations and encourage further research in this area, aiming to deepen our understanding of ASR's role in enhancing writing skills and promoting effective language learning.

References

Ahn, T., & Lee, S. (2016). User experience of a mobile speaking application with automatic speech recognition for EFL learning. *British Journal of Educational Technology*, *47*(4), 778–786. https://doi.org/10.1111/bjet.12354

Annamalai, N., & Tan, K. E. (2015). Exploring two teachers' engagement with their students in an online writing environment. *The EuroCALL Review*, *23*(2), 58–73. https://doi.org/10.4995/EUROCALL.2015.4668

Arcon, N., Klein, P. D., & Dombroski, J. D. (2017). Effects of dictation, speech to text, and handwriting on the written composition of elementary school English language learners. *Http://Dx.Doi.Org.Lib-Ezproxy.Concordia.ca/10.1080/10573569.2016.1253513*, *33*(6), 533–548. https://doi.org/10.1080/10573569.2016.1253513

Ardasheva, Y., & Tretter, T. R. (2013). Strategy inventory for language learning-ELL student form: Testing for factorial validity. *Modern Language Journal*, *97*(2), 474–489. https://doi.org/10.1111/j.1540-4781.2013.12011.x

Baron, N. S. (2000). *Alphabet to email: How written English evolved and where it's heading*. Routledge Kegan Paul.

Behizadeh, N., & Engelhard, G. (2011). Historical view of the influences of measurement and writing theories on the practice of writing assessment in the United States. *Assessing Writing*, *16*(3), 189–211. https://doi.org/10.1016/j.asw.2011.03.001

Berninger, V. W., & Amtmann, D. (2003). Preventing written expression disabilities through early and continuing assessment and intervention for handwriting and/or spelling problems: Research into practice. In H. L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 345–363). Guilford Press.

Berninger, V. W., Fuller, F., & Whitaker, D. (1996). A process model of writing development across the life span. *Educational Psychology Review*, *8*(3), 193–218. https://www.jstor.org/stable/23359413

Berninger, V. W., Garcia, N. P., & Abbott, R. D. (2009). Multiple processes that matter in writing instruction and assessment. In *Instruction and assessment for struggling writers: Evidence-based practices* (pp. 15–50). Guilford Press.

Berninger, V., & Winn, W. (2006). Implications of advancements in brain research and technology for writing development, writing instruction, and educational evolution. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 96–114). Guilford Press.

Berninger, V., Yates, C., Cartwright, A., Rutberg, J., Remy, E., & Abbott, R. (1992). Lower-level developmental skills in beginning writing. *Reading and Writing*, *4*(3), 257–280. https://doi.org/10.1007/BF01027151

Bodnar, S., Cucchiarini, C., Penning de Vries, B., Strik, H., & van Hout, R. (2017a). Learner affect in computerised L2 oral grammar practice with corrective feedback. *Computer Assisted Language Learning*, *30*(3–4), 223–246. https://doi.org/10.1080/09588221.2017.1302964

Bodnar, S., Cucchiarini, C., Penning de Vries, B., Strik, H., & van Hout, R. (2017b). Learner affect in computerised L2 oral grammar practice with corrective feedback. *Computer Assisted Language Learning*, *30*(3–4), 223–246. https://doi.org/10.1080/09588221.2017.1302964

Bolter, J. D. (2001). *Writing space: Computers, hypertext, and the remediation of print* (2nd ed.). Routledge.

Bourdin, B. (2000). Is graphic activity cognitively costly? A developmental approach. *Reading and Writing*, *13*, 183–196. https://doi.org/10.1023/A:1026458102685

Bourdin, B., & Fayol, M. (1994). Is written language production more difficult than oral language production? A working memory approach. In *International Journal of Psychology* (Vol. 29, Issue 5). https://doi.org/10.1080/00207599408248175

Breetvelt, I., Van den Bergh, H., & Rijlaarsdam, G. (1996). Rereading and generating and their relation to text quality. An application of multilevel analysis on writing process data. In G. Rijlaarsdam, H. van den Bergh, & M. Couzijn (Eds.), *Theories, Models and Methodology in Writing Research* (pp. 10–20). Amsterdam University Press.

British Council. (n.d.). *LearnEnglish*. Retrieved March 14, 2022, from https://learnenglish.britishcouncil.org/skills/reading

Burston, J. (2003). Software Selection: A Primer on Sources and Evaluation. *CALICO Journal*, *21*(1), 29–40. https://doi.org/10.1558/cj.v21i1.29-40

Cardoso, W. (2022). Technology for Speaking Development. In T. M. Derwing, M. J. Munro, & R. I. Thomson (Eds.), *Routledge Handbook on Second Language Acquisition and Speaking*. Routledge. https://www.routledge.com/The-Routledge-Handbook-of-Second-Language-Acquisition-and-Speaking/Derwing-Munro Thomson/p/book/9780367900847

Caute, A., & Woolf, C. (2016). Using voice recognition software to improve communicative writing and social participation in an individual with severe acquired dysgraphia: an experimental single-case therapy study. *Aphasiology*, *30*(2–3), 245–268. https://doi.org/10.1080/02687038.2015.1041095

Cavalier, A., & Ferretti, R. (1996). Talking instead of typing: Alternate access to computers via speech recognition technology. *Focus on Autism and Other Developmental Disabilities*, *11*(2), 79–85. https://doi.org/10.1177/108835769601100203

Chamot, A. U. (2005). Language learning strategy instruction: Current issues and research. *Annual Review of Applied Linguistics*, *25*, 112–130. https://doi.org/10.1017/S0267190505000061

Chamot, A. U. (2007). Accelerating academic achievement of English language learners. In J. Cummins & C. Davison (Eds.), *International handbook of English language teaching* (pp. 317–331). Springer. https://doi.org/10.1007/978-0-387-46301-8_23%0A%0A

Chapelle, C. (2001). *Computer Applications in Second Language Acquisition: Foundations for Teaching, Testing, and Research*. Cambridge University Press.

Chapelle, C., & Jamieson, J. (2008). *Tips for teaching with CALL: Practical approaches to computer-assisted language learning*. Pearson Education.

Chaudron, C. (1988). Learner behavior in second language classrooms. *Second Language Classrooms*, 90–117. https://doi.org/10.1017/CBO9781139524469.006

Chen, H. (2011). Developing and evaluating Synctolearn, a fully automatic video and transcript synchronization tool for EFL learners. *Computer Assisted Language Learning*, *24*(2), 117–130. https://doi.org/10.1080/09588221.2010.526947

Chen, H. H. J. (2011). Developing and evaluating an oral skills training website supported by automatic speech recognition technology. *ReCALL*, *23*(1), 59–78. https://doi.org/10.1017/S0958344010000285

Chiu, T., Liou, H., & Yeh, Y. (2007). A Study of web-based oral activities enhanced by Automatic Speech Recognition for EFL college learning. *Computer Assisted Language Learning*, *20*(3), 209–233. https://doi.org/10.1080/09588220701489374

Chukharev-Hudilainen, E., & Klepikova, T. A. (2016). The effectiveness of computer-based spaced repetition in foreign language vocabulary instruction: A double-blind study. *CALICO Journal*, *33*(3), 334–354. https://doi.org/10.1558/cj.v33i3.26055

Clark, R. (1983). Reconsidering research on learning from media. *Review of Educational Research*, *53*(4), 445–459. https://doi.org/10.3102/00346543053004445

Cohen, A. D. (2014). *Strategies in learning and using a second language*. Routledge.

Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science*, *1*(3), 98–101. https://doi.org/10.1111/1467-8721.EP10768783/ASSET/1467-8721.EP10768783.FP.PNG_V03

Covington, M. A., & McFall, J. D. (2010). Cutting the Gordian Knot: The Moving-Average Type-Token Ratio (MATTR). *Journal of Quantitative Linguistics*, *17*(2), 94–100. https://doi.org/10.1080/09296171003643098

Cox, T., & Davies, R. (2012). Using automatic speech recognition technology with elicited oral response testing. *CALICO Journal*, *29*(4), 601–618. https://www.jstor.org/stable/10.2307/calicojournal.29.4.601

Cox, T. L., & Davies, R. S. (2012). Using Automatic Speech Recognition Technology with Elicited Oral Response Testing. *CALICO Journal*, *29*(4), 601–618. https://doi.org/10.11139/CJ.29.4.601-618

Crowhurst, M. (1980). Syntactic complexity in narration and argument at three grade levels. *Canadian Journal of Education*, *5*(1), 6–13. https://www.jstor.org/stable/1494634

Cucchiarini, C., Nejjari, W., & Strik, H. (2012). My pronunciation coach: improving English pronunciation with an automatic coach that listens. *Language Learning in Higher Education*, *1*(2), 365–376. https://doi.org/10.1515/cercles-2011-0024

Dalby, J., & Kewley-Port, D. (2013). Explicit Pronunciation Training Using Automatic Speech Recognition Technology. *CALICO Journal*, *16*(3), 425–445. https://doi.org/10.1558/CJ.V16I3.425-445)

Daniels, P., & Iwago, K. (2017). The suitability of cloud-based speech recognition engines for language learning. *JALT CALL Journal*, *13*(3), 229–239. https://doi.org/10.29140/jaltcall.v13n3.220

Darwin, C. (1871). *The Descent of Man and Selection in Relation to Sex*. Murray.

Davis, F. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly: Management Information Systems*, *13*(3), 319–339. https://doi.org/10.2307/249008

Davis, F., Bagozzi, R., & Warshaw, P. (1992). Extrinsic and Intrinsic Motivation to Use Computers in the Workplace. *Journal of Applied Social Psychology*, *22*(14), 1111–1132. https://doi.org/10.1111/j.1559-1816.1992.tb00945.x

Davis, S. (2019, September 12). *13 Thought-Provoking Personal Narrative Prompts*. Academic Writing Success. https://www.academicwritingsuccess.com/13-thought-provoking-personal-narrative-prompts/

de Vries, B., Cucchiarini, C., Bodnar, S., Strik, H., & van Hout, R. (2015). Spoken grammar practice and feedback in an ASR-based CALL system. *Computer Assisted Language Learning*, *28*(6), 550–576. https://doi.org/10.1080/09588221.2014.889713

Dekeyser, R. (1993). The effect of error correction on L2 grammar knowledge and oral proficiency. *The Modern Language Journal*, *77*(4), 501–514. http://dx.doi.org/10. 1111/ j.1540-4781.1993.tb01999.x

DeVellis, R. (2003). *Scale development: Theory and applications* (2nd ed.). Sage Publications.

Dillon, T., & Wells, D. (2021). Student Perceptions of Mobile Automated Speech Recognition for Pronunciation Study and Testing. *English Teaching(South Korea)*, *76*(4), 101–122. https://doi.org/10.15858/engtea.76.4.202112.101

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Routledge.

Drijbooms, E., Groen, M., & Verhoeven, L. (2017). How executive functions predict development in syntactic complexity of narrative writing in the upper elementary

grades. *Reading and Writing*, *30*(1), 209–231. https://doi.org/10.1007/s11145-016-9670-8

Egbert, J., & Shahrokni, S. (2018). *CALL Principles and Practices*. Open Educational Resources (OER).

Elimat, A., & AbuSeileek, A. (2014). Automatic speech recognition technology as an effective means for teaching pronunciation. *The JALT CALL Journal*, *10*(1), 21–47. https://doi.org/10.29140/jaltcall.v10n1.166

Enge, E. (2020). *Mobile Voice Usage Trends in 2020*. Perficient. https://www.perficient.com/insights/research-hub/voice-usage-trends

Ertmer, P. A., Ottenbreit-Leftwich, A. T., Sadik, O., Sendurur, E., & Sendurur, P. (2012). Teacher beliefs and technology integration practices: A critical relationship. *Computers & Education*, *59*(2), 423–435. https://doi.org/10.1016/J.COMPEDU.2012.02.001

Evers, K., & Chen, S. (2021). Effects of automatic speech recognition software on pronunciation for adults with different learning styles. *Journal of Educational Computing Research*, *59*(4), 669–685. https://doi.org/10.1177/0735633120972011

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Addison-Wesley.

Flower, L., & Hayes, J. (1980). The cognition of discovery: Defining a rhetorical problem. *College Composition and Communication*, *31*(1), 21–32. https://doi.org/10.2307/356630

Flower, L., & Hayes, J. (1981). A cognitive process theory of writing. *College Composition and Communication*, *32*(4), 365–387. https://www.jstor.org/stable/356600?seq=1&cid=pdf-reference#references_tab_contents

Fox, C., Israelsen-Augenstein, M., Jones, S., & Gillam, S. (2021). An evaluation of expedited transcription methods for school-age children's narrative language: Automatic speech

recognition and real-time transcription. *Journal of Speech, Language, and Hearing Research*, *64*(9), 3533–3548. https://doi.org/10.1044/2021_JSLHR-21-00096

Garvin, P. L., & Ladefoged, P. (1963). Speaker Identification and Message Identification in Speech Recognition. *Phonetica*, *9*(4), 193–199. https://doi.org/10.1159/000258404/MACHINEREADABLECITATION/RIS

Gass, S. (1997). *Input, interaction, and the second language learner*. Lawrence Erlbaum Associates.

Geertz, C. (2008). Thick Description: Toward an interpretive theory of culture. In *The cultural geography reader* (1st ed., pp. 41–51). Routledge.

Gil Urrutia, J. I., Brangier, E., & Cessat, L. (2017). Is a holistic criteria-based approach possible in user experience? In A. Marcus & W. Wang (Eds.), *Design, user experience, and usability: Theory, methodology, and management* (pp. 395–409). Springer, Cham. https://doi.org/10.1007/978-3-319-58634-2_29

Graham, S. (1990). The role of production factors in learning disabled students' compositions. *Journal of Educational Psychology*, *82*(4), 781–791. https://doi.org/10.1037/0022-0663.82.4.781

Grant, M., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information and Libraries Journal*, *26*(2), 91–108. https://doi.org/10.1111/j.1471-1842.2009.00848.x

Grenfell, M., & Harris, V. (1999). *Modern languages and learning strategies* (1st ed.). Routledge. https://doi.org/10.4324/9780203013823

Griffiths, C. (2007). Language learning strategies: Students' and teachers' perceptions. *ELT Journal*, *61*(2), 91–99. https://doi.org/10.1093/elt/ccm001

Guiraud, P. L. (1960). *Problèmes et méthodes de la statistique linguistique* (1st ed.). Springer Dordrecht. https://link.springer.com/book/9789027700254

Haas, C. (1987). What research with computers can tell us about the uses of reading in writing. *Paper Presented at the 1987 Conference on College Composition and Communication*.

Haas, C. (1989). Seeing it on the screen isn't really seeing it: Computer writers' reading problems. In G. E. Hawisher & C. L. Selfe (Eds.), *Critical Perspectives on Computers* (pp. 16–29). Teachers College Press.

Hartley, J., Sotto, E., & Pennebaker, J. (2003). Speaking versus typing: A case-study of the effects of using voice-recognition software on academic correspondence. *British Journal of Educational Technology*, *34*(1), 5–16. https://doi.org/10.1111/1467-8535.00300

Hartwick, J., & Barki, H. (1994). Explaining the role of user participation in information system use. *Management Science*, *40*(4), 440–465. https://doi.org/10.1287/mnsc.40.4.440

Hayes, J. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 1–27). Lawrence Erlbaum Associates.

Healey, B., & Gardner, P. (2021). Writing, grammar and the body: a cognitive stylistics framework for teaching upper primary narrative writing. *Literacy*, *55*(2), 125–135. https://doi.org/10.1111/lit.12242

Honeycutt, L. (2003). Researching the use of voice recognition writing software. *Computers and Composition*, *20*(1), 77–95. https://doi.org/10.1016/S8755-4615(02)00174-3

Honeycutt, L. (2004). Literacy and the writing voice: The intersection of culture and technology in dictation. *Journal of Business and Technical Communication*, *18*(3), 294–327. https://doi.org/10.1177/1050651904264105

Hsu, L. (2016). An empirical examination of EFL learners' perceptual learning styles and acceptance of ASR-based computer-assisted pronunciation training. *Computer Assisted Language Learning*, *29*(5), 881–900. https://doi.org/10.1080/09588221.2015.1069747

Hubbard, P. (1988). An integrated framework for CALL courseware evaluation. *CALICO Journal*, *6*(2), 51–72. https://doi.org/10.1558/cj.v6i2.51-72

Hubbard, P. (1996). Elements of CALL methodology: Development, evaluation, and implementation. In M. Pennington (Ed.), *The power of CALL* (Vol. 96, pp. 15–32). Athelstan.

Hubbard, P. (2006). Evaluating CALL Software. In L. Ducate & N. Arnold (Eds.), *Calling on CALL: From theory and research to new directions in foreign language teaching* (pp. 313–338). CALICO.

IELTS Guide. (2022). Cambridge Assessment English. https://www.ielts.org/

Inceoglu, S., Chen, W. H., & Lim, H. (2023). Assessment of L2 intelligibility: Comparing L1 listeners and automatic speech recognition. *ReCALL*, *35*(1), 89–104. https://doi.org/10.1017/S0958344022000192

Isaac, S., & Michael, W. (1995). *Handbook in research and evaluation: A collection of principles, methods, and strategies useful in the planning, design, and evaluation of studies in education and the behavioral sciences* (3rd ed.). EdITS Publishers.

Jiang, M., Jong, M., Lau, W., Chai, C., & Wu, N. (2021). Using automatic speech recognition technology to enhance EFL learners' oral language complexity in a flipped classroom. *Australasian Journal of Educational Technology*, *37*(2), 110–131. https://doi.org/10.14742/AJET.6798

Johnson, R. B., & Onwuegbuzie, A. J. (2004). *Mixed Methods Research: A Research Paradigm Whose Time Has Come*. *33*(7), 14–26. http://dx.doi.org/10.3102/0013189X033007014

Johnson, R., & Christensen, L. (2019). *Educational research: Quantitative, qualitative, and mixed approaches* (7th ed.). Sage Publications. https://bookshelf.vitalsource.com/#/books/9781544337821/

Julious, S. A. (2005). Sample size of 12 per group rule of thumb for a pilot study. *Pharmaceutical Statistics*, *4*(4), 287–291. https://doi.org/10.1002/PST.185

Jurafsky, D., & Martin, J. (2000). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall.

Kellogg, R. (1996). A model of working memory in writing. In C. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 57–71). Lawrence Erlbaum Associates.

Kellogg, R. (2001). Competition for working memory among writing processes. *American Journal of Psychology*, *114*(2), 175–191. http://dx.doi.org/10.2307/1423513

Kyle, K., Crossley, S. A., & Jarvis, S. (2021). Assessing the Validity of Lexical Diversity Indices Using Direct Judgements. *Language Assessment Quarterly*, *18*(2), 154–170. https://doi.org/10.1080/15434303.2020.1844205

Lackey, N., & Wingate, A. (1998). The pilot study: One key to research success. In P. J. Brink & M. J. Wood (Eds.), *Advanced design in nursing research* (2nd ed.). Sage.

Lan, R., & Oxford, R. L. (2003). Language learning strategy profiles of elementary school students in Taiwan. *IRAL - International Review of Applied Linguistics in Language Teaching*, *41*(4), 339–379. https://doi.org/10.1515/iral.2003.016

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, *16*(3), 307–322. https://doi.org/10.1093/applin/16.3.307

Lavrakas, P. (2008). Intercoder reliability. In *Encyclopedia of Survey Research Methods*. Sage Publications. https://dx.doi.org/10.4135/9781412963947.n228

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

Leijten, M., & Van Waes, L. (2005). Writing with speech recognition: The adaptation process of professional writers with and without dictating experience. *Interacting with Computers*, *17*(6), 736–772. https://doi.org/10.1016/j.intcom.2005.01.005

Leijten, M., & Van Waes, L. (2006). Repair strategies in writing with speech recognition: The effect of experience with classical dictating. In L. Van Waes, M. Leijten, & C. Neuwirth (Eds.), *Writing and Digital Media* (pp. 31–46). Brill. https://doi.org/10.1163/9781849508209_005

Liakin, D., Cardoso, W., & Liakina, N. (2015). Learning L2 pronunciation with a mobile speech recognizer: French /y/. *CALICO Journal*, *32*(1), 1–25. https://doi.org/10.1558/cj.v32i1.25962

Liakin, D., Cardoso, W., & Liakina, N. (2017). Mobilizing Instruction in a Second-Language Context: Learners' Perceptions of Two Speech Technologies. *Languages*, *2*(3), 11. https://doi.org/10.3390/languages2030011

Linnarud, M. (1986). Lexis in composition: A performance analysis of Swedish learners' written English. *Studies in Second Language Acquisition*, *9*(2), 254–256. https://doi.org/10.1017/S0272263100000516

Long, M. (1983). Linguistic and conversational adjustments to non-native speakers. *Studies in Second Language Acquisition*, *5*(2), 177–193. https://www.jstor.org/stable/44487277

Ma, R., & Oxford, R. L. (2014). A diary study focusing on listening and speaking: The evolving interaction of learning styles and learning strategies in a motivated, advanced ESL learner. *System*, *43*(1), 101–113. https://doi.org/10.1016/j.system.2013.12.010

Macaro, E., Handley, Z., & Walter, C. (2012). A systematic review of CALL in English as a second language: Focus on primary and secondary education. *Language Teaching*, *45*(1), 1–43. https://doi.org/10.1017/S0261444811000395

MacArthur, C. (1999). Overcoming barriers to writing: Computer support for basic writing skills. *Reading and Writing Quarterly*, *15*(2), 169–192. https://doi.org/10.1080/105735699278251

MacArthur, C. (2000). New tools for writing: Assistive technology for students with writing difficulties. *Topics in Language Disorders*, *20*(4), 85–100.

MacArthur, C., & Cavalier, A. (2004). Dictation and speech recognition technology as test accommodations. *Exceptional Children*, *71*(1), 43–58. https://doi.org/10.1177/001440290407100103

Magogwe, J. M., & Oliver, R. (2007). The relationship between language learning strategies, proficiency, age and self-efficacy beliefs: A study of language learners in Botswana. *System*, *35*(3), 338–352.

Matsuhashi, A. (1982). Explorations in real-time production of written discourse. In M. Nystrand (Ed.), *What Writers Know: The Language, Process, and Structure of Written Discourse* (pp. 269–290). Academic Press.

Mayer, R. E. (2005). Cognitive theory of multimedia learning. *The Cambridge Handbook of Multimedia Learning.*

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, *42*(2), 381–392. https://doi.org/10.3758/BRM.42.2.381/METRICS

McCrocklin, S. (2016). Pronunciation learner autonomy: The potential of automatic speech recognition. *System*, *57*, 25–42.

McCrocklin, S. (2019). Learners' feedback regarding ASR-based dictation practice for pronunciation learning. *CALICO Journal*, *36*(2), 119–137. https://doi.org/10.1558/cj.34738

McCrocklin, S., & Edalatishams, I. (2020). Revisiting Popular Speech Recognition Software for ESL Speech. *TESOL Quarterly*, *54*(4), 1086–1097. https://doi.org/10.1002/tesq.3006

McCrocklin, S., Humaidan, A., & Edalatishams, E. (2019). ASR dictation program accuracy: Have current programs improved? In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th pronunciation in second language learning and teaching conference* (pp. 191–200). Iowa State University.

McCutchen, D. (1996). A capacity theory of writing: Working memory in composition. *Educational Psychology Review*, *8*(3), 299–325. https://doi.org/10.1007/BF01464076

McCutchen, D., Covill, A., Hoyne, S., & Mildes, K. (1994). Individual differences in writing: Implications of translating fluency. *Journal of Educational Psychology*, *86*(2), 256–266. https://psycnet.apa.org/doi/10.1037/0022-0663.86.2.256

McNamara, D., Crossley, S., & McCarthy, P. (2010). Linguistic features of writing quality. *Written Communication*, *27*(1), 57–86. https://doi.org/10.1177/0741088309351547

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press.

Mehdipour-Kolour, D., Khaksaran, K., & Bilal Ali, M. (2018). Vygotskianism and the Enhancement of Mobile Language Learnerss Argumentative Writing Skill. *SSRN Electronic Journal*, 26–41. https://ssrn.com/abstract=3178990

Miller, G., Eugene, G., & Pribram, K. (1960). *Plans and the Structure of Behaviour*. Routledge.

Mirzaei, M., Meshgi, K., Akita, Y., & Kawahara, T. (2017). Partial and synchronized captioning: a new tool to assist learners in developing second language listening skill. *ReCALL*, *29*(2), 178–199. https://doi.org/10.1017/S0958344017000039

Moats, L., & Tolman, C. . (2009). *Speaking is natural: Reading and writing are not.* Reading Rockets. http://www.readingrockets.org/article/28758?theme=print

Moore, G., & Benbasat, I. (1991). Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information Systems Research*, *2*(3), 192–222. https://doi.org/10.1287/isre.2.3.192

Moussalli, S., & Cardoso, W. (2019). Intelligent personal assistants: can they understand and be understood by accented L2 learners? *Computer Assisted Language Learning*, *0*(0), 1–26. https://doi.org/10.1080/09588221.2019.1595664

Mroz, A. (2018a). Seeing how people hear you: French learners experiencing intelligibility through automatic speech recognition. *Foreign Language Annals*, *51*(3), 617–637. https://doi.org/10.1111/flan.12348

Mroz, A. (2018b). Seeing how people hear you: French learners experiencing intelligibility through automatic speech recognition. *Foreign Language Annals*, *51*(3), 617–637. https://doi.org/10.1111/flan.12348

Nation, I. S. P., & Newton, J. M. (2008). *Teaching ESL/EFL Listening and Speaking*. Taylor & Francis.

Nguyen, T., Hwang, W., Pham, X., & Zhao-heng, M. (2018). User-oriented EFL speaking through application and exercise: instant speech translation and shadowing in authentic context. *Journal of Educational Technology & Society*, *21*(4), 129–142. https://www.jstor.org/stable/10.2307/26511544

Oxford, R. L. (1986). *Development and psychometric testing of the Strategy Inventory for Language Learning (SILL)*.

Oxford, R. L. (1999). Relationship between second language strategies and proficiency in the context of learner autonomy and self-regulation. *Revista Canaria de Estudios Ingleses [Canadian Journal of English Studies]*, *1*(38), 109–126.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu\, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Systematic Reviews*, *10*(89). https://doi.org/10.1186/s13643-021-01626-4

Pawlak, M., & Kruk, M. (2022). *Individual differences in computer assisted language learning research*. https://doi.org/10.4324/9781003240051

Pérez Castillejo, S. (2021). Automatic speech recognition: Can you understand me? In T. Beaven & F. Rosell-Aguilar (Eds.), *Innovative language pedagogy report* (pp. 121–126). Research-publishing.net. https://doi.org/10.14705/rpnet.2021.50.1246

Perreault, C., & Mathew, S. (2012). Dating the origin of language using phonemic diversity. *PloS One*, *7*(4), 1–8. https://doi.org/10.1371/journal.pone.0035289

Porter, D. (1960). A report on instructional devices in foreign language teaching. In A. Lumsdaine & R. Glaser (Eds.), *Teaching machines and programmed learning: A source bookRT ON INSTRUCTIONAL DEVICES IN FOREIGN LANGUAGE TEACHING.* (pp. 186–205). National Education Associtaion.

Powell, B. (2009). *Writing: Theory and history of the technology of civilization*. John Wiley & Sons.

Quinlan, T. (2004). Speech recognition technology and students with writing difficulties: Improving fluency. *Journal of Educational Psychology*, *96*(2), 337–346. https://doi.org/10.1037/0022-0663.96.2.337

Quinlan, T. (2006a). Young writers and digital scribes. In *Writing and digital media* (pp. 21–30). Brill.

Quinlan, T. (2006b). Young writers and digital scribes. In V. L. Waes, M. Leijten, & C. Neuwirth (Eds.), *Writing and Digital Media* (pp. 21–29). Brill. https://brill.com/view/title/23238

Reece, J., & Cummings, G. (1996). Evaluating speech-based composition methods: Planning, dictation, and the listening word processor. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 361–380). Lawrence Erlbaum Associates.

Richards, J. (2015). *Key issues in language teaching*. Cambridge University Press.

Rosen, K., & Yampolsky, S. (2000). Automatic speech recognition and a review of its functioning with dysarthric speech. *Augmentative and Alternative Communication*, *16*(1), 48–60.

Saadatzi, M., Pennington, R., Welch, K., Graham, J., & Scott, R. (2017). The use of an autonomous pedagogical agent and automatic speech recognition for teaching sight words to students with autism spectrum disorder. *Journal of Special Education Technology*, *32*(3), 173–183. https://doi.org/10.1177/0162643417715751

Selouani, S., Lê, T., Benahmed, Y., & O'Shaughnessy, D. (2008). Speech-enabled tools for augmented interaction in e-learning applications. *International Journal of Distance Education Technologies*, *6*(2), 1–20. https://doi.org/10.4018/jdet.2008040101

Sidgi, L., & Shaari, A. (2017). The Effect of Automatic Speech Recognition EyeSpeak Software on Iraqi Students' English Pronunciation: A Pilot Study. *Advances in Language and Literary Studies*, *8*(2), 48. https://doi.org/10.7575/aiac.alls.v.8n.2p.48

Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing: The ESL research andits implications. *TESOL Quarterly*, *27*, 657–677. https://doi.org/10.2307/3587400

Silva, W., Carchedi, L., Junior, J., de Souza, J., Barrere, E., & de Souza, J. (2021). A framework for large-scale automatic fluency assessment. *International Journal of Distance Education Technologies*, *19*(3), 70–88. https://doi.org/10.4018/IJDET.2021070105

Smith, E., Sumner, P., Hedge, C., & Powell, G. (2021). Smart speaker devices can improve speech intelligibility in adults with intellectual disability. *International Journal of Language & Communication Disorders*, *56*(3), 583–593. https://doi.org/10.1111/1460-6984.12615

Sohail, M. (2020). Exploring user satisfaction with AI-enabled voice-activated smart phone assistants. *PACIS*. https://aisel.aisnet.org/pacis2020

Stewner-Manzanares, G., Chamot, A. U., O'Malley, J. M., Kupper, L., & Russo, R. P. (1984). *A teacher's guide for using learning strategies in English-as-a-second-language instruction*.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass & C. G. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Newbury House.

Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning* (1st ed., pp. 471–483). Routledge. https://doi.org/10.4324/9781410612700

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1016/0364-0213(88)90023-7

Tynjälä, P., Mason, L., & Lonka, K. (2001). Writing as a learning tool: An introduction. In P. Tynjälä, L. Mason, & K. Lonka (Eds.), *Writing as a learning tool: Integrating theory and practice* (Vol. 7, pp. 7–22). Springer, Dordrecht. https://doi.org/10.1007/978-94-010-0740-5_2

Van Belle, G. (2002). *Statistical rules of thumb*. John Wiley.

van Doremalen, J., Boves, L., Colpaert, J., Cucchiarini, C., & Strik, H. (2016). Evaluating automatic speech recognition-based language learning systems: a case study. *Computer Assisted Language Learning*, *29*(4), 833–851. https://doi.org/10.1080/09588221.2016.1167090

Van Lieshout, C., & Cardoso, W. (2022). Google Translate as a Tool for Self-Directed Language Learning. *Language Learning & Technology*, *26*(1), 1–19. http://hdl.handle.net/10125/73460

Venkatesh, V., & Davis, F. (2000a). Theoretical extension of the Technology Acceptance Model: Four longitudinal field studies. *Management Science*, *46*(2), 186–204. https://doi.org/10.1287/mnsc.46.2.186.11926

Venkatesh, V., & Davis, F. D. (2000b). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, *46*(2), 186–204.

Walker, N. R., Cedergren, H., Trofimovich, P., & Gatbonton, E. (2011). Automatic speech recognition for CALL: A task-specific application for training nurses. *Canadian Modern Language Review*, *67*(4), 459–479. https://doi.org/10.3138/cmlr.67.4.459

Wang, Y., & Young, S. (2015). Effectiveness of feedback for enhancing English pronunciation in an ASR-based CALL system. *Journal of Computer Assisted Learning*, *31*(6), 493–504. https://doi.org/10.1111/jcal.12079

Weigle, S. (2010). Assessing Writing. In *Cambridge University Press*. Cambridge University Press. https://doi.org/https://doi.org/10.1017/CBO9780511732997

Wetzel, K. (1996). Speech-recognizing computers: A written-communication tool for students with learning disabilities? *Journal of Learning Disabilities*, *29*(4), 371–380.

Wood, E., Willoughby, T., Specht, J., & Porter, L. (2002). An examination of how a cross-section of academics use computer technology when writing academic papers. *Computers and Education*, *38*, 287–301.

Young, S., & Wang, Y. (2014). The game embedded CALL system to facilitate English vocabulary acquisition and pronunciation. *International Forum of Educational Technology & Society*, *17*(3), 239–251. https://www.jstor.org/stable/10.2307/jeductechsoci.17.3.239

Young, V., & Mihailidis, A. (2010). Difficulties in automatic speech recognition of dysarthric speakers and implications for speech-based applications used by the elderly: A literature review. *Assistive Technology*, *22*(2), 99–112. https://doi.org/10.1080/10400435.2010.483646

Yu, D., & Deng, L. (2014). *Automatic Speech Recognition: A Deep Learning Approach*. Springer. https://doi.org/10.1007/978-1-4471-5779-3

Yu, G. (2010). Lexical diversity in writing and speaking task performances. *Applied Linguistics*, *31*(2), 236–259. https://doi.org/10.1093/applin/amp024

Yu, P., Pan, Y., Li, C., Zhang, Z., Shi, Q., Chu, W., Liu, M., & Zhu, Z. (2016). User-centred design for Chinese-oriented spoken english learning system. *Computer Assisted Language Learning*, *29*(5), 984–999. https://doi.org/10.1080/09588221.2015.1121877

Zenker, F., & Kyle, K. (2021). Investigating minimum text lengths for lexical diversity indices. *Assessing Writing*, *47*, 100505. https://doi.org/10.1016/j.asw.2020.100505

Appendices

Appendix A: Consent Form

INFORMATION AND CONSENT TO PARTICIPATE IN A RESEARCH STUDY

Study Title: Exploring Automatic Speech Recognition as a Writing Tool

Researcher: Danial Mehdipour Kolour
Researcher's Contact Information: danial.mehdipourkolour@concordia.ca
Faculty Supervisor: Dr. Walcir Cardoso
Faculty Supervisor's Contact Information: 514-848-2424 ext. 2451/ walcir.cardoso@concordia.ca
Source of funding for the study: Dr. Walcir Cardoso

You are being invited to participate in the research study mentioned above. This form provides information about what participating would mean. Please read it carefully before deciding if you want to participate or not. If there is anything you do not understand, or if you want more information, please ask the researcher.

A. PURPOSE

The purpose of the research is to investigate the benefits of using automatic speech recognition (ASR) technology for writing purposes.

B. PROCEDURES

If you participate, you will be asked to complete a number of activities as follows:

- Schedule a Zoom meeting with the researcher, using Doodle.
- Sign this consent letter.
- Practice using ASR to write. You will be asked to write two tasks: one free writing (with no writing prompt – 5 minutes) and one with a visual writing prompt (5 minutes). Training will be provided by the researcher. Your practice and training will NOT be recorded.
- To check your proficiency: complete a dictation activity in which you read a text out loud to the ASR in Google Docs. Your voice and your screen will be recorded (5-10 minutes).
- Complete two writing tasks using Google Doc's ASR. Your audio and screen will be recorded (30-40 minutes).
- Fill out a survey about your experience writing with ASR (about 20 minutes).
- (Optional) Participate in an interview of about 20 minutes; it will be conducted at your most convenient time. The interview will be audio recorded.

In total, your participation including the interview will be approximately 80 minutes.

C. RISKS AND BENEFITS

There will be no known risks to the participants of this research.

This research is not intended to benefit you personally, but if you participate, you will learn how to use a voice-based technology for writing. You will also receive some tips to improve your writing skills.

D. CONFIDENTIALITY

We will gather the following information as part of this research:

- Your basic demographic information such as age, and your perceptions of using voice to write – ASR – through a survey and interview
- We will record your audio and screen activities while doing the dictation activity and writing tasks; your image will not be included in the recording.
- Nobody will have access to the collected data, except two people directly connected to this research: my supervisor, Dr. Walcir Cardoso, and me. The data will only be used for the research described in this form.
- The information gathered will be coded (your name will be removed), meaning that your real identity will be kept confidential, so it will not be possible to make a link between you and the information you provide.
- We will protect the information by allowing only two people who are directly connected to this research (my supervisor, Dr. Walcir Cardoso, and me) to access the digital data, including the online survey, dictation output, writing responses, and interview transcripts, and recordings. All the collected digital data will be stored in the researcher's password protected computer.
- We intend to publish the results of the research as part of my PhD dissertation. However, it will not be possible to identify you in the published results. One year after I complete my PhD, all the data collected from you will be destroyed.

F. CONDITIONS OF PARTICIPATION

You do not have to participate in this research. It is purely your decision. If you do participate, you can stop at any time. You can also ask that the information you provided not be used, and your choice will be respected. If you decide that you don't want us to use your information, you must tell the researcher before December 31st, 2021.

- As a compensatory indemnity for participating in this research, you will receive $10 for the first set of Zoom activities (initial training, dictation activity, writing tasks, and online survey), and then another $10 compensation for participating in an optional interview.
- To make sure that research money is being spent properly, auditors from Concordia or outside will have access to a coded list of participants. It will not be possible to identify you from this list.
- There are no negative consequences for not participating, stopping in the middle, or asking us not to use your information.

G. PARTICIPANT'S DECLARATION

I have read and understood this form. I have had the chance to ask questions and any questions have been answered. I agree to participate in this research under the conditions described.

NAME (please print) _____

SIGNATURE _____

DATE _____

If you have questions about the scientific or scholarly aspects of this research, please contact the researcher. Their contact information is on page 1. You may also contact their faculty supervisor.

If you have concerns about ethical issues in this research, please contact the Manager, Research Ethics, Concordia University, 514.848.2424 ex. 7481 or oor.ethics@concordia.ca.

Appendix B: Background Questionnaire

This background questionnaire collects a range of information about your demographic profile, preferred mode of transcription, and language background.

Email:

Demographic Profile

1. What is your age range? 20-24 ( ) – 25-29 ( ) – 30-34 ( ) – 35 years or older ( )
2. What is your gender? Male ( ) Female ( ) Prefer not to say ( ) Other ( )
3. What academic degree are you currently pursuing?
    1. Undergraduate 2. Graduate

Transcription Mode Preference

4. What do you often use to write in English?

    Mobile device ☐ Computer ☐ Paper ☐

5. Have you ever tried practicing any language skills using automatic speech recognition (ASR)? [ASR is voice-to-text technology that allows you to use your voice to type; this can be found in Google Docs].
    Yes ☐ No ☐
    1. If so, which skill(s)/subskills(s)? _____
    2. What specific was it? _____

6. Do you use your voice when texting someone?
    Yes ☐ No ☐

Language Background

7. How long have you been leaning English? (if you are non-native speaker of English)
    _____

8. What is (are) your first language(s)? Indicate if there is more than one.
    1. _____
    2. _____

9. Did you first learn to write in you first language(s)? If no, indicate the language(s) in which you first learned to write. Yes No
    1. _____

2. _____
3. _____

10. What language(s) can you speak? Write down from most proficient to least. You may start with your first language.
    1. _____ (your first language)
    2. _____
    3. _____

11. In what language(s) can you read? Write down from most proficient to least. You may start with your first language.
    4. _____
    5. _____
    6. _____

12. In what language(s) can you write? Write down from most proficient to least. You may start with your first language.
    7. _____
    8. _____
    9. _____

13. What are the first two language skills (e.g., speaking, writing, listening, reading) you find most difficult to learn?
    10. _____
    11. _____

14. How long have you lived in Canada (in years)?

    _____

Appendix C: Free Writing Prompt

Training Phase:

For practice, using Google's Voice Typing, compose a piece of writing on any subject of your choice within a time frame of 5 to 10 minutes.

Appendix D: Visual Writing Prompt

Training Phase:

Look at the image below and use your voice to write down any ideas that come to your mind within 5 to 10 minutes.

Appendix E: Dictation Activity

Training Phase:

You may read the following texts out loud to ASR. This part will be recorded.

Level B1

Every time you're online, you are bombarded by pictures, articles, links and videos trying to tell their story. Unfortunately, not all of these stories are true. Sometimes they want you to click on another story or advertisement at their own site, other times they want to upset people for political reasons. These days it's so easy to share information. These stories circulate quickly, and the result is fake news.
There is a range of fake news: from crazy stories which people easily recognise to more subtle types of misinformation. Experts in media studies and online psychology have been examining the fake news phenomenon.
Look at the website where the story comes from. Does it look real? Is the text well written? Are there a variety of other stories or is it just one story? Fake news websites often use addresses that sound like real newspapers, but don't have many real stories about other topics. If you aren't sure, click on the 'About' page and look for a clear description.

Level B2

If we look around us at the things we have purchased at some point in our lives, we would no doubt notice that not everything we own is being put to good use: the thick woolen coat which we thought looked trendy despite the fact that we live in a tropical country, the smartphone that got put away when we bought ourselves the newest model, the car that only gets used at the weekends, or even the guest room in our house that somehow got turned into a storeroom.
Those underutilised items may seem useless to some, but could be an asset to others. With the advent of the internet, online communities have figured out a way to generate profit from the sharing of those underused assets. Using websites and social media groups that facilitate the buying and selling of second-hand goods, it is now easier than ever for peer-to-peer sharing activities to take place. And this is known as the sharing economy.

Appendix F: Narrative Writing Tasks

Choose **TWO** of the three following narrative writing prompts to voice write your narrative. You may spend about **20 minutes** on each narrative.

1. First Love
First loves are powerful and vivid stories. Consider a story about how you met that person or a significant memory between you. This could be something wonderful, funny, or even terrible, if the relationship is over how did it end? Were you heartbroken? These stories, or a story about any important relationship, leave an imprint on us.

2. **Success Story**
Write a story about something you accomplished. Describe what you did to achieve success, and how you reached your goal. Were there obstacles you had to overcome? This is a tricky essay topic because sometimes people focus on the feelings of success or the final achievement rather than the story of how they succeeded. Some ideas for this topic are to write about winning an event, earning an honor, confronting a personal challenge, or getting over a phobia.

3. **Personal Failure**
Write about how you failed to do or failed to achieve something. Describe what you tried to do, how you failed, and what you gained/learned from it. Everyone fails at something. Your readers understand the feeling of failure. Some people connect with these personal narratives more than stories of success because they empathize with you. People learn a lot through failure, so when you write this type of story, hint or convey what you learned from failing.

Appendix G: Online User Perception Survey

You are invited to participate in our survey about your perception of using speech recognition technology (ASR) for writing purposes (i.e., voice-inputted text). It will take around 20 minutes to complete this survey.

Your participation in this survey is completely voluntary and there are no risks associated with this project. However, if you feel uncomfortable answering any questions, you can skip the question or withdraw from the survey at any point.

Your responses to the background questionnaire and user perception survey will be strictly con131dential. The data from this research will be coded and reported only in general terms (nobody will be able to identify you in these reports). If you have questions at any time about the survey or the procedures, you may contact Danial Mehdipour Kolour by email at danial.mehdipourkolour@concordia.ca

For each of the questions below choose the numeric response that best characterizes your opinion about the statement.

Construct	Measurement statement
Perceived Ease of Use (PEU)	- I find ASR easy to use. (PEU_1) - Learning how to create voice-inputted text is easy. (PEU_2) - It is easy to become skillful at producing voice-inputted text. (PEU_3) - I find it easy to get ASR to write what I want to write. (PEU_4) - Interacting with ASR does not require a lot of mental effort. (PEU_5) - Interacting with ASR is clear and understandable. (PEU_6)
Perceived Usefulness (PU)	- Using ASR improves my writing performance. (PU_1) - Using ASR increases my writing productivity. (PU_2) - Using ASR enhances my effectiveness in writing because it removes transcription challenges such as spelling. (PU_3) - Creating voice-inputted text enhances my performance in formal writing such as essays or research papers. (PU_4) - I find ASR to be useful in increasing my writing quality. (PU_5) - Using ASR helps me to produce more ideas (i.e., brainstorming) when writing. (PU_6)
Intention to Use (IU)	- Assuming that I have access to the Internet, I intend to use ASR as a writing tool. (IU_1) - Assuming that I have access to ASR, I predict that I would use it for writing purposes. (IU_2) - Assuming that I have access to ASR, I would rather create voice-inputted text than typed text. (IU_3) - I intend to be a heavy user of ASR. (Heavy user: someone who uses a system frequently. (IU_4)
Attitude (AT)	- I am positive towards using ASR as a writing tool. (AT_1) - It is a good idea to use ASR for writing purposes. (AT_2) - It is a practical decision to create voice-inputted texts, as opposed to typed texts. (AT_3)

Subjective Norm (SN)	- People who influence my behavior think that I should use ASR as a writing tool. (SN_1) - As a university student, it is important for me to use ASR for writing purposes. (SN_2) - I should use ASR because of the similarity of my values and society values in its use. (SN_3) - For me to be ready for future writing tasks, I find it necessary to use ASR. (SN_4)
Output Quality (OQ)	- The quality of ASR output (i.e., voice-inputted text) is high. (OQ_1) - As a user, I experienced no problem with the quality of ASR's output. (OQ_2) - The quality of ASR output does not frustrate the process of writing. (OQ_3)
Result Demonstrability (RD)	- I believe I could share my experience of using ASR with others. (RD_1) - I would find it difficult to explain why creating voice-inputted text may or may not be useful. (RD_2) - I would find it easy to communicate to others the results of using ASR. (RD_3)

The questions of this survey are adapted from the *Technology Acceptance Model 2* (TAM2; Venkatesh & Davis, 2000) and Park (2009).

Note: All items were measured on a 5-point Likert scale (conducted online):
1 = Strongly agree, 2 = Agree, 3 = Neutral, 4 = Disagree, 5 = Strongly Disagree

For example:
I find ASR easy to use.
1 = Strongly agree | 2 = Agree | 3 = Neutral | 4 = Disagree | 5 = Strongly Disagree

Appendix H: Pilot Study Results

Construct	Measurement statement	M	SD	*r*	Cronbach's α
Perceived Ease of Use (PEU)	- I find ASR easy to use. (PEU_1)	2.00	1.12	.73	.89
	- Learning how to create voice-inputted text is easy. (PEU_2)	2.00	1.20	.66	
	- It is easy to become skillful at producing voice-inputted text. (PEU_3)	2.25	1.13	.86	
	- I find it easy to get ASR to write what I want to write. (PEU_4)	2.50	1.38	.85	
	- Interacting with ASR does not require a lot of mental effort. (PEU_5)	2.33	1.15	.72	
	- Interacting with ASR is clear and understandable. (PEU_6)	1.67	.65	.55	
Perceived Usefulness (PU)	- Using ASR improves my writing performance. (PU_1)	3.00	1.12	.95	.93
	- Using ASR increases my writing productivity. (PU_2)	2.29	1.37	.94	
	- Using ASR enhances my effectiveness in writing because it removes transcription challenges such as spelling. (PU_3)	2.50	1.38	.46	
	- Creating voice-inputted text enhances my performance in formal writing such as essays or research papers. (PU_4)	3.17	1.33	.74	
	- I find ASR to be useful in increasing my writing quality. (PU_5)	2.92	1.31	.94	
	- Using ASR helps me to produce more ideas (i.e., brainstorming) when writing. (PU_6)	2.92	1.16	.83	
Intention to Use (IU)	- Assuming that I have access to the Internet, I intend to use ASR as a writing tool. (IU_1)	2.92	1.31	.85	.93
	- Assuming that I have access to ASR, I predict that I would use it for writing purposes. (IU_2)	2.75	1.35	.90	
	- Assuming that I have access to ASR, I would rather create voice-inputted text than typed text. (IU_3)	3.33	1.43	.88	
	- I intend to be a heavy user of ASR. (Heavy user: someone who uses a system frequently. (IU_4)	3.33	1.61	.76	
Attitude (AT)	- I am positive towards using ASR as a writing tool. (AT_1)	2.25	1.13	.70	.79
	- It is a good idea to use ASR for writing purposes. (AT_2)	2.25	.96	.43	

	- It is a practical decision to create voice-inputted texts, as opposed to typed texts. (AT_3)	2.75	1.13	.80	
Subjective Norm (SN)	- People who influence my behavior think that I should use ASR as a writing tool. (SN_1)	3.83	1.15	.59	.74
	*- As a university student, it is important for me to use ASR for writing purposes. (SN_2)	3.33	1.23	.33	
	*- I should use ASR because of the similarity of my values and society values in its use. (SN_3)	3.00	1.34	.17	
	- For me to be ready for future writing tasks, I find it necessary to use ASR. (SN_4)	3.92	1.08	.59	
Output Quality (OQ)	- The quality of ASR output (i.e., voice-inputted text) is high. (OQ_1)	2.50	1.08	.65	.83
	- As a user, I experienced no problem with the quality of ASR's output. (OQ_2)	3.58	1.08	.70	
	- The quality of ASR output does not frustrate the process of writing. (OQ_3)	3.25	1.13	.73	
Result Demonstrability (RD)	- I believe I could share my experience of using ASR with others. (RD_1)	2.17	1.33	.76	.89
	*- I would find it difficult to explain why creating voice-inputted text may or may not be useful. (RD_2)	2.08	1.08	.31	
	- I would find it easy to communicate to others the results of using ASR. (RD_3)	2.33	.98	.69	
Total		2.69	1.19	.69	.85

Appendix I: Demographic Characteristics of Participants

Variable	Category	*N*	*%*
Age			
	20-24	18	30.0
	25-29	12	20.0
	30-34	14	23.0
	35 or older	16	27.0
Gender			
	Male	22	36.7
	Female	37	61.7
	Prefer not to say	1	1.7
Level of education			
	Undergraduate	22	36.7
	Graduate	38	63.3
Language			
	NS of English	30	50.0
	NNS of English	30	50.0
First language(s)			
	Arabic	3	5.0
	Bengali	2	3.4
	Cantonese	2	3.4
	Chinese	2	3.4
	Farsi	13	22.0
	French	6	10.0
	German	2	3.4
	Hindi	2	3.4
	Portuguese	1	1.7
	Spanish	1	1.7
	Turkish	1	1.7
	Taiwanese	1	1.7
	Vietnamese	3	5.0
Preferred tool for writing			
	Computer	46	76.7
	Smartphone (or other type of mobile devices)	8	13.3
	Pen and paper	6	10.0
Prior use of ASR for language practice			
	No	44	73.3
	Yes	16	26.7
NNS use of ASR for language skills			
	Speaking	7	41.3
	Pronunciation	6	35.4
	Writing	2	11.8
Use of ASR for general purposes			
	Checking weather forecast	2	3.4
	Checking distance	1	1.7
	Note taking	2	3.4

	Text messaging	8	13.3
Number of years of living in an English-speaking country			
	1-3	11	18.3
	4-6	5	8.3
	More than 6	9	15.0
	Never	7	11.7