# Exploring Intelligent Personal Assistants in Second Language Acquisition

Souheila Moussalli

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

Montréal, Québec, Canada

May 2022

CONCORDIA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Souheila Moussalli

Entitled: Exploring Intelligent Personal Assistants in Second Language Acquisition

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:

_____ Chair
Dr. Mary Horst

_____ Thesis Supervisor
Dr. Walcir Cardoso

_____ Thesis Supervisor
Dr. Sara Kennedy

_____ Examiner
Dr. Saul Carliner

_____ Examiner
Dr. Denis Liakin

_____ External Examiner
Dr. Marie-Josee Hamel

Approved by _____
Dr. Steven Shaw, Graduate Program Director

May 2022 _____
Dr. Pascale Sicotte, Dean of Arts and Sciences

## Abstract

Exploring Intelligent Personal Assistants in Second Language Acquisition

Souheila Moussalli, Ph.D.
Concordia University, 2022

The goal of this dissertation is to investigate Intelligent Personal Assistants (IPAs), a voice-controlled service that can complete various functions by orally interacting with its users, as pedagogical tools in English second language classrooms to assess their pedagogical suitability. This dissertation begins with a review of the literature focusing on the importance of using technology in the language classroom. The remainder is divided into three manuscript-based chapters in which each manuscript addresses one aspect of the general research questions: (a) What are language learners' perceptions of the use of IPAs as learning tools? (Manuscript A); (b) Can IPAs understand different language learners, and can these learners understand IPAs? (Manuscript B); and (c) Can IPAs help English language learners improve their receptive and productive skills? (Manuscript C).

The first manuscript investigates the use of IPAs and users' perceptions of the technology as a language learning tool. It examines a number of variables such as the IPAs' ease of use, options for learner self-regulation (defined as learners' ability to understand and control their learning environment), learner motivation and, more importantly, opportunities for learner input and output practice. The second manuscript explores IPA's ability to interact with different accented language learners of English. The focus is on exploring the IPA's ability to understand speech from different levels of language accentedness, and vice versa: to explore learners' ability to understand the synthesized speech. The third manuscript investigates whether the pedagogical use of IPAs can lead to improvements in learners' phonological awareness, perception and

production of the allomorphy that characterizes regular past tense -ed marking in English (example depending on the preceding phonological environment, suffix -ed can be pronounced as talk/t/, play/d/ and add/id/).

This dissertation contributes to our knowledge of learner experience and attitudes towards IPAs as it can further unfold the potentials and limitations of the technology. As far as second language phonology/pronunciation is concerned, the dissertation breaks new ground in research since little is known about IPAs and their pedagogical potential for the development of second language listening and speaking skills.

**Acknowledgements**

I would like to thank both my supervisors, Dr. W. Cardoso and Dr. S. Kennedy. My sincerest thanks and appreciation go to Dr. Walcir Cardoso, who was abundantly helpful and offered invaluable assistance, support, and guidance writing this dissertation and during my tenure at Concordia University. Thank you for your dedication, encouragement, and counsel. You are always available and ready to provide whatever support your students seek or need. All this embedded in an ever-smiling personality! I will be forever grateful. You are truly an inspiration, and I consider myself very fortunate to have been mentored by one of the best.

Another heartfelt gratitude and words of appreciation also go to my dear supervisor Dr. Sarita Kennedy, for her encouragement, guidance and advice which have been invaluable to me throughout my studies. I am glad to have had my first research experience with you. Working under your supervision has been very enjoyable and I have learned a lot from you.

I would also like to thank my committee members, Dr. Saul Carliner and Dr. Denis Liakin for their valuable insights and assistance. Thank you for all your feedback. Your input continuously had a positive impact on my work, and I will be forever grateful for the role you played in the development of my dissertation. I would also like to thank Dr. Marie-Josee Hamel for being part of my Examination Committee, and to Dr. Mary Horst for chairing the defense.

To my husband, thank you for the constant support, countless sacrifices, and endless love when it was most needed. You have made me stronger, better, and more fulfilled than I could have ever imagined. I definitely could not have completed this project without your encouragement. To my lovely son and daughter, thank you for your patience and understanding and for being the best kids ever! You are my inspiration to achieve greatness!

Last, but not least, I would like to express my gratitude to my family members – my parents and siblings – for their support, blessings and wishes for the successful completion of this project. I hope that I have made you proud.

## Contribution of Authors

This manuscript-based dissertation consists of three manuscripts, along with a general introduction (Chapter 1) that states the research rationale and objectives and a concluding chapter (Chapter 5). The two published articles (Chapters 2 and 3) will be presented in their entirety, as stated in Concordia's 2020 edition of 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 overlap in the content between manuscripts.

This manuscript-based dissertation was conceptualized over the course of my studies and meetings with my supervisors, Dr. Walcir Cardoso and Dr. Sara Kennedy, and in the Computer Assisted Language Learning (CALL) research group meetings in the Department of Education at Concordia University. As the first author for the included three manuscripts, I was the major contributor to the manuscripts' conception, content, design, data collection, and write-up, under the guidance of Dr. Cardoso. This contribution is reflected in my status as first author in all three manuscripts.

## Contents

List of Figures ..... x
List of Tables ..... x
List of Definitions ..... xi
Chapter One: General Introduction ..... 1
Introduction to IPAs ..... 4
An IPA Example: Amazon Echo and Alexa ..... 7
Why IPAs and Echo ..... 9
Investigating IPAs in Computer Assisted Language Learning ..... 11
IPAs and Second Language Education ..... 12
The Current Dissertation ..... 15
Chapter Two: Practicing English with a speaking robot: Learners' perceptions ..... 18
Introduction and Background ..... 18
TTS and ASR in the language classroom ..... 20
Method ..... 23
Participants ..... 23
Amazon Echo and Alexa ..... 24
Procedure ..... 25
Analysis ..... 26
Results ..... 26
Survey ..... 26
Interview ..... 27
Discussion and Conclusion ..... 28
Chapter Three: Intelligent personal assistants: ..... 32
Can they understand and be understood by accented L2 learners? ..... 32
Intelligent Personal Assistants and Technology-Enhanced L2 Education ..... 33
The Target Smart Speaker and IPA: Amazon Echo and Its Voice, Alexa ..... 37
IPA-Human Interaction and Accented Speech ..... 39
Human-human Interaction and Accented Speech ..... 39
Human-machine Interactions ..... 42
The Current Study ..... 44
Method ..... 45

Instruments ..... 45
Questionnaire ..... 45
Surveys 46
Participants' ratings and transcriptions of the IPA's speech ..... 46
Interview 47
Judges' ratings and transcriptions of participants' speech ..... 47
Treatment ..... 48
Analysis ..... 49
Results ..... 51
Discussion and concluding remarks ..... 54
Limitations and Future Directions ..... 57
Chapter Four: Intelligent Personal Assistants in L2 Education: Focus on English Past -ed ..... 59
IPAs and Alexa ..... 61
Attitudes towards the use of the Alexa App: Student perceptions ..... 64
Simple Past Tense -ed in English ..... 65
The Communicative Framework for Teaching Pronunciation ..... 67
Participants ..... 69
Material and Design ..... 70
Instruments ..... 70
Raters ..... 73
Procedure ..... 75
Analysis ..... 78
Results ..... 79
Limitations and Future Directions ..... 92
References ..... 107
Appendices ..... 120
Appendix A: Consent Form ..... 120
Appendix B: Consent Form ..... 121
Appendix C: Language Background Questionnaire ..... 124
Appendix D: ECHO/ ALEXA Target questions ..... 126
Appendix E: Survey ..... 128
Appendix F: Consent Form ..... 129
Appendix G: Assessment of Echo's speech (by L2 leaners) ..... 132

Appendix H: Assessment of Echo's speech (by judges) ..... 133
Appendix I: Consent Form ..... 134
Appendix J: Understanding the simple past tense ..... 136
Appendix K: pre-test and post- test perception words ..... 137
Appendix L: pre-test and post- test perception sentences ..... 138
Appendix M: pre-test production words ..... 141
Appendix N: pre-test controlled mock interview ..... 142
Appendix O: Survey Learner's perception of Alexa App as a pedagogical tool ..... 145
Appendix P: practice the –ed forms ..... 148
Appendix Q: Interview questions ..... 149
Appendix R: Practice with Alexa -ed sounds ..... 150

## List of Figures

Figure 1. 1 ..... 6
Figure 1.2 ..... 8
Figure 1.3 ..... 8
Figure 2.1 ..... 24
Figure 2.2 ..... 25
Figure 4.1 ..... 62
Figure 4.2 ..... 63
Figure 4.3 ..... 71
Figure 4.4 ..... 74

## List of Tables

Table 2. 1 ..... 26
Table 3.1 ..... 51
Table 3.2 ..... 52
Table 3.3 ..... 52
Table 3.4 ..... 53
Table 3.5 ..... 54
Table 4. 1 ..... 76
Table 4.2 ..... 79
Table 4.3 ..... 80
Table 4.4 ..... 82
Table 4.5 ..... 82
Table 4.6 ..... 83
Table 4.7 ..... 84
Table 4.8 ..... 85

## List of Definitions

**Affordances**: It is the relationship between a user and the object being used, as to what a user can do with the object based on the user's capabilities (Gibson, 1979). In this dissertation, the term refers to how language learners interact and use IPAs for pedagogical purposes, mainly to practice aural and oral skills.

**Accented Speech**: Following Derwing and Munro (2009), it is 'the way in which speech differs from the local variety of [that speech] and the impact of that difference on speakers and listeners' (p. 476). The concept of accentedness includes two sub-components: intelligibility and comprehensibility, discussed below.

**Alexa listening mode**: A feature that allows Alexa to interact with its users by listening to them as they speak in order to respond to the oral request and perform the required task.

**Alexa Skills:** According to Amazon, skills are like apps for Alexa that provide a new channel for user content and services. Skills allow users to use their voices to perform everyday tasks and services like checking the weather, playing a game, and more. Generally, users can create and publish these skills in the Alexa skill store to reach other users.

**Comprehensibility**: "The listener's perception of how easy or difficult it is to understand a given speech sample" (Derwing & Munro, 2009 p. 479). In this dissertation, comprehensibility is measured via listeners' perceptions of what they heard, often assessed using a Likert-type rating scale, ranging from "extremely difficult to understand" to "extremely easy to understand".

**Computer Assisted Language Learning.** "[T]he area of applied linguistics concerned with the use of computers for teaching and learning a second language" (Chapelle, 2008) in which a "learner uses a computer and, as a result, improves his or her language" (Beatty, 2003).

**Intelligibility**: "The degree of a listener's actual comprehension of an utterance" (Derwing & Munro, 2009, p. 478), often measured via transcriptions of what is produced by the L2 speaker under consideration.

**Motivation:** According to Dörnyei (2009) second language (L2) motivation consists of three dimensions: 1. Ideal L2 Self, a powerful motivator to learn the second language to become a competent L2 speaker through reducing the discrepancy between actual and ideal selves; 2. Ought-to Self, which refers to the possible self or 'outcome self' that the leaner wants to achieve; and 3. L2 Learning Experience, which concerns the immediate learning context and language learner experience. This definition aligns with Gardner's (1985) view, adopted in this dissertation. For the author, for learners to be motivated, they need to have something to look forward to, a purpose related to a goal or objective. In this study, IPAs are hypothesized to promote these objectives, as they engage learners in human-machine interactions and problem solving.

**Pedagogical Effectiveness**: It is a broad area of inquiry that includes pedagogical affect and course effectiveness (Clarke & Nelson, 2012). In this dissertation, the term refers to the process of evaluating the performance of IPAs as they are being used by the learners to practice oral and

aural skills. This evaluation is concerned with the attitudes and emotions of learners towards the use of IPAs as they interact with the IPAs using a specific set of knowledge, skill, and technique.

**Second Language Skills:** The four basic language skills are speaking, listening, reading, and writing. In this dissertation, the focus is on listening and speaking skills.

**Usability:** According to Baven (1995), usability is the ease of use and acceptability of a product for a particular class of users carrying out specific tasks in a specific environment. In this dissertation, usability refers to IPAs' ability to achieve specific goals with effectiveness, efficiency, and satisfaction by learners. For details about the importance of this theoretical construct, see Lah, Lewis, and Šumak (2020).

## Chapter One: General Introduction

Technological devices have become an integral part of our daily lives. In contrast to the past, today many people feel comfortable using laptops, phones, iPads, and related software in many areas of their lives, as they have provided people in general and learners in specific with time, flexibility, and access to a wide range of knowledge and information. Their importance has been accentuated further with the rise of the global pandemic the world is facing, in which learners have become digital learners by necessity, and the process of education, the content of education, and its pedagogical practices have been profoundly impacted. Our views of education and learning shifted in ways we have never imagined before. For instance, today, learners' real-life experiences are now combined with virtual experiences in which lessons are adapted and/or customized to meet their learning goals and accommodate their new learning environment (Kinshuk & Graf, 2012). In addition, during this health crisis, learners have been able to interact with content, with other learners, with their teachers and different technological tools anywhere, anytime and at their own pace (Chapelle, 2003; Driscoll & Carliner, 2005). This recurrent user-driven and self-and-group initiated practice has led learners into "learning in the wild" (Sauro & Zourou, 2019), which is an informal, dynamic, and unpredictable learning environment which occurs in digital spaces beyond the context of formal instruction and in an out-of-class context. This type of learner-technology practice is believed to enhance learners' overall pedagogical experiences (Nunan & Richards, 2014).

The use of technology is thus seen as a supportive tool for teachers and an effective approach to learning (for a detailed review and similar claims, see Kukulska-Hulme, 2012 for Mobile assisted language learning; Chapelle, 2003 for Computer assisted language learning; Lee, 2001 & Rouhshad et al. 2015 for Computer mediated communication; and Alastuey, 2012, and

Fernandez-Garcia & Martinez Arbelaiz, 2002, 2003 for asynchronous and/or synchronous computer-mediated communication). Furthermore, learners have shown positive perceptions toward the use of technology in their language classrooms (Peters et al., 2009). A study by Taylor and Gitsaki (2004), for instance, showed that students benefited from using the web in their English language class and they considered it as a rewarding tool for language learning. These findings thus suggest that, when properly used, modern technology can be a valuable add-on to face-to-face language teaching.

Moreover, particularly in human-machine environments, special attention is placed on learners' interaction and negotiation of meaning, which are vital elements of second language acquisition (Samani et al., 2015). Long's (2983; 1996) Interaction Hypothesis attempts to explain language learning through learners' exposure to the target language (input), learners' production of the target language (output), and feedback on that production. As we know it today, this hypothesis conjectures that for effective and efficient language learning to occur, four constructs need to be targeted: comprehensible input, comprehensible output, feedback, and negotiation of meaning. According to Long (1996), opportunities for learning a language develop when interlocutors engage in a conversation using the target language and a misunderstanding arises. In that scenario, both speakers and listeners have to overcome the misunderstanding by negotiating for meaning. The same applies for oral human-machine interactions, which are considered helpful in promoting oral production and conversation skills. The benefit of human-machine interactions is the environment itself, which supplies rich input, promotes output, provides plentiful and dynamic feedback, focuses learners' attention to aspects and features of the language being learned, and enhances noticing (another important aspect of second langauge learning).

As explained earlier, input and output are vital for language development but in many language classrooms, learners are faced with many challenges. Some of these challenges include a lack of access to different varieties of target language input (e.g., aural content) and time constraints, which results in limited opportunities to practice and interaction with other learners and/or teachers (Al-Hosni, 2014; Collins et al., 2009; Gaparini, 2005; Lopes, 2007). To complicate the matter, during class time, many students feel anxious to communicate in the target language for fear of losing face (Neri et al., 2003) or because they deem their speaking abilities inadequate (Nazara, 2011). As a result, students avoid using the second language and, consequently, the quantity (and possibly the quality) of teacher- or peer-based feedback is also reduced.

To address these challenges, both learners and teachers can benefit from the use of modern technologies in the classroom. For instance, learners can rely on them for listening and speaking practice in terms of quality (e.g., a desired dialectal variety, a focus on certain pronunciation features) and quantity (the amount of target language exposure). Language teachers can use them to foster and add great value to their teaching in a variety of contexts (Felix, 2001). An example of such technology is voice-controlled intelligent personal assistants (IPAs henceforth; see discussion in the next section) which are emergent worldware, meaning they are technology not developed for educational purposes but repurposed for that use. I believe that the pedagogical use of IPAs can be a valuable addition to second language teaching because these devices have the potential to enhance both the quantity and quality of input that learners receive (e.g., by increasing access to second language and making some aspects of the input salient - see Cardoso, 2018; Collins et al., 2009). In addition, they can motivate students to practice their oral skills in the target language and provide additional opportunities for communication outside of the classroom in a

stress free, meaningful, fun environment, anytime and anywhere. Consequently, these affordances- which is defined as what a user can do with an object based on the user's capabilities- may contribute to an increase in the learners' willingness to communicate using the target language (Gregersen & MacIntyre, 2014; Moussalli & Cardoso, 2016). Finally, the use of IPAs has the potential to promote learner autonomy, as the technology may expand learning opportunities to out-of-class contexts and, at the same time, foster a personalized, learner centered, and collaborative second language pedagogy (Kim & Kwon, 2012).

Certainly, for the adoption of new tools such as IPAs, it is important to assess their pedagogical suitability (e.g., by examining the technology's affordances using Chapelle's 2001 criteria for selecting Computer Assisted Language Learning (CALL) tasks), understand how learners perceive them, and whether they can lead learners to learning gains. Consequently, this dissertation aims to investigate IPAs as supportive tools for teachers and students that can extend the reach of the classroom, promote learner interaction, and enhance and aid the language learning process. This dissertation starts with a general introduction (the current chapter) that introduces IPAs, explains why IPAs were chosen, and the method adopted for investigating IPAs within CALL research. It is then followed by three main manuscripts that address the following three topics: Learners' perceptions of IPAs (Chapter 2); appropriateness of IPAs for second language pedagogical purposes (e.g., can the technology understand and be understood by its users? – Chapter 3), and the tool's ability to contribute to learning (e.g., can IPAs contribute to the learning of a specific second language feature – the pronunciation of English past -ed? – Chapter 4).

**Introduction to IPAs**

Intelligent Personal Assistants (or IPA) are voice-controlled services that can perform various tasks by interacting with its users using natural language user interface (de Barcelos Silva

et al., 2020). To function and perform properly, IPAs use a combination of three technologies: a conversational interface, personal context awareness, and service delegation. The conversational interface relies on automatic speech recognition (ASR) tools (i.e., voice recognition, voice analysis and language processing) to interpret speech and consequently meaning. The second technology, personal context awareness, enables IPAs to understand their users' language patterns and specific nuance, like a person's variation in tone (e.g., for questions) or using words that usually occur in similar syntactic positions (e.g., noun phrases such as "Mary" in utterance-initial position, indicating the start of a new sentence). Finally, service delegation allows IPAs to integrate with other programs and apps owned by the users. For example, users can connect their IPAs to other smart home apps or electrical appliances or websites to perform certain tasks (e.g., to turn their lights on). Both conversational interface and personal context awareness allow IPAs to communicate with users in "human-like conversations" (Johnson et al., 2014) because of the ASR capability which permits users to speak rather than type in information.

ASR helps users by transmitting and reciting speech or by generating captions of conversations such as discussions and lectures. For example, in the context of voice-activated web searches, based on a person's oral request for a location (e.g., 'Where is Montreal located?'), the built-in ASR system takes advantage of cloud computing by transforming that oral request into text, which is then used to look for answers in the cloud or a text-based database (i.e., the web). After the information requested is located, the application's Text-To-Speech (TTS) synthesizer outputs the answer in the form of computer-generated speech, thus making the user hear an oral reply to the original request. Some of the most popular IPAs that use the above functions include Amazon Alexa, Google home, and Apple Siri. It is worth mentioning as well that these IPAs also work as apps on devices other than the ones manufactured by the company

that publishes the software – some of the most popular IPAs are illustrated in Figure1.1.

It should be noted that the mentioned intelligent personal assistants, whose main function is to simulate human-like conversations, were developed earlier throughout the mid-1900s and they were given different names as their functions evolved. Their function evolved from text-to-text interaction as in the case of chatbots (software application that converse with their users by providing answers to their questions via text messages on chat due to natural language processes) such as ELIZA (see Weizenbaum, 1966). ELIZA, one of the first chatbots, analysed input sentences and created responses based on reassembly rules associated with a decomposition of the input. Yet, ELIZA could not enter any form of sophisticated interaction skills because it could not hold any memory of the conversation (Kerly et al., 2007). Since then, significant development in syntactic language processing led to the development of speech-to-text, text-to-speech and to a combination of functions, as in the case of artificial intelligent speech synthesizers, and virtual assistants. A systematic review of published IPA research showed that most research has focused on infrastructure and usability, but "domains such as education […] are unexplored" (de Barcelos Silva et al., 2020).

In this dissertation, the name *speaking robot* was initially used in the first manuscript to refer to Amazon's Echo. However, in the second and third manuscripts, the term evolved to "smart speaker" or IPAs to reflect the combination of functionality of a speaker and an intelligent assistant. In the next section, Alexa, which is the voice assistant associated to the smart speaker Amazon Echo (and the one adopted in this research) will be presented.

**Figure 1. 1**

*Examples of IPA*


**An IPA Example: Amazon Echo and Alexa**

As illustrated in Figure 1.2, Amazon Echo is a tall cylinder (9.25-inch/23.5 cm) hands-free speaker that connects to its associated app, the cloud-based voice service Alexa, which provides prompt oral answers to any questions asked. Its modus operandi resembles that of other voice-controlled personal assistants like Google Home, Apple Siri and Microsoft Cortana. In addition to its attributes as a conversational partner, the Alexa App, an IPA associated with Echo, provides transcriptions of the interaction between the user and Echo. It also allows users to enable different categories of skills that range from education, gaming and social life to business (see Figure 1.3 for how users can browse the Alexa App to look for certain categories or skills). When Echo detects the wake word (i.e., "Alexa", or the phrase that causes the device, Echo, to begin listening to a user's request so that it can be processed), it lights up and streams audio to the cloud, where the Alexa voice service recognizes and responds to the request. Another interesting feature is that Echo is speaker adaptive: The more a person uses Echo, the more its

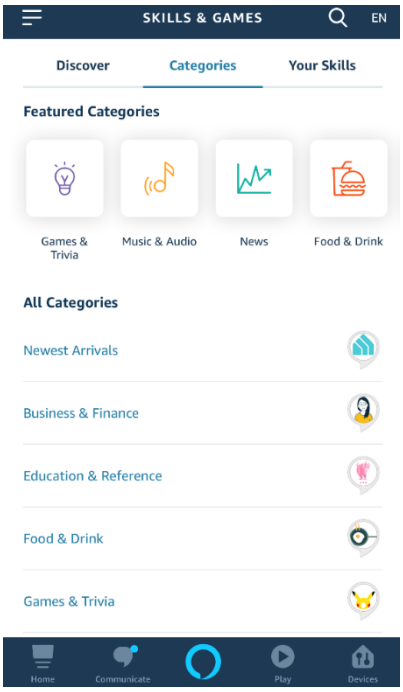
voice software Alexa adapts to the speech patterns, vocabulary, and the user's personal preferences. However, this adaptation or accuracy improvement sparks privacy concern issues because of the amount of data that is being gathered from users, mainly by listening to their recordings and learning so much about them

**Figure 1.2**

*Amazon Echo*


**Figure 1.3**

*Alexa App*




## Why IPAs and Echo

The present dissertation adopts a commercial and easily available smart speaker, Amazon Echo, and its associated IPA App, Alexa, for a variety of reasons. First, Echo is widely accessible, less costly than other similar devices, and its Alexa App is free to download (the app can be used by itself without a smart speaker such as Echo). Second, the technology has third-party skills and features which are enabled freely. These add value to the IPA as being a cost-effective tool to be used by teachers and learners with little or no funding. Nevertheless, teachers and learners should be cautious when using these skills as they create potential security and privacy concerns. These concerns could range from misleading privacy policies to the ability of third-parties to change their program codes after receiving Amazon approval, leading to potential leakage of sensitive personal information (Lentzsch et al., 2021). Third, Echo was rated as the best "smart speaker" in 2020 (Segan & Greenwald 2020) and 2018 (Segan, 2018), and its associated voice Alexa was considered the most popular IPA in 2020 (i.e., it was downloaded

nearly 10 million times; Forbes; Valerie 202; Reviews.com staff, 2020) and in 2017 (with a 70% share of the voice-controlled speaker market; Rao, 2017). Lastly, keeping in mind that Echo and its associated Alexa App are worldware (i.e., technology not developed for educational purposes but repurposed for that use), they fulfil Chapelle's requirements for adopting CALL resources, which will be discussed next.

Chapelle (2001) brought forth seven criteria to assess a tool's suitability for adoption in second language pedagogy:

(1) Reliability and learner fit: Is the use of the technology user friendly (not too difficult to use)? Do learners' individual differences (e.g., their learning style, motivation, and experience with technology) affect their performance?

(2) Authenticity and generalizability: How does the use of the technology reflect technology use in the non-research environment? Can research results be generalized to other contexts?

(3) Construct validity and operationalization of learning conditions: What theoretical constructs underlie the use of technology? How does the use of technology reflect those constructs?

(4) Language learning potential and operationalization of learning conditions: What potential does the use of technology have for language learning? How is that potential realized in its use in research?

(5) Interactiveness and meaning focus: How does the use of technology engage learners' meaningful use of communicative abilities?

(6) Positive impact: How do learners benefit or suffer from use of the technology?

(7) Practicality: How easy is it to find, modify, and use the technology in a non-research

context?

This dissertation seeks to demonstrate that IPAs such as Amazon's Echo/Alexa meet these criteria. For example, they can be adapted to meet the learners' own interests and needs (learner fit). Their use is authentic and can be generalized to other contexts (authenticity and generalizability) and they have great potential to promote learning via increased input and output practice (language learning potential). In addition, they can have a positive impact on learners' autonomy and general well-being (positive impact) and they are easy to use in out-of-the-classroom contexts since learners only require the device and a natural human skill: their voices (practicality). Finally, IPAs foster human-machine interactions that can be meaning focused (meaningful interactions), and the pedagogy involved can be motivated by many theoretical constructs (e.g., Gardner & Miller's, Self-Access Language Learning, 1999; Long 's Interactionist Approach, 1996; and Wong's input flood, 2005).

**Investigating IPAs in Computer Assisted Language Learning**

Given that IPAs are fairly new and very little is known about their pedagogical use, it is important to examine their pedagogical suitability. Cardoso (2022) explains that CALL research traditionally involves four general stages for examining the pedagogical potential of new technologies (such as the IPA targeted in this dissertation). These stages include:

1- Development: focuses on the development of a new tool. This stage is not relevant to this dissertation because IPAs have already been developed and they are widely used, as discussed earlier.

2- Exploration: involves examining the pedagogical potential and affordance of new technology. In general, at this stage, researchers motivate the use of new technologies by looking at how they meet a set of practical and/or theoretical criteria for effective

pedagogy – see the discussion around Chapelle (2001) above (Chapters 1 and 2 address this stage, directly or indirectly).

3- Assessing suitability: involves research on usability, acceptance, and learner's attitudes towards the technology, which can be investigated through different quantitative, qualitative or mixed methods (Chapter 2 and 3 address this stage).

4- Assessing pedagogical effectiveness: involves assessing the pedagogical effectiveness of the new technology. At this stage, research focuses on the technology's ability to promote learning (e.g., learning gains, which is the focus of Chapter 4).

Following Cardoso' (2022) chronological framework, the current dissertation examines stages 2 through 4 by exploring IPAs as tools for second language pedagogy, with each chapter focusing on a particular (but sometimes overlapping) stage. For example, as indicated above, chapters 2 and 3 investigate stages 2 and 3, where the focus is on examining learners' perceptions of IPAs and the tool's appropriateness as pedagogical tools. Chapter 4, on the other hand, addresses stage 4 by examining IPAs' potential to promote learning.

To motivate and contextualize the scope of the proposed dissertation, the next section provides a literature review of IPAs in second language education.

**IPAs and Second Language Education**

As indicated above, typical CALL research tends to follow a chronology that starts with the development of a technology, then focuses on its suitability for pedagogical purposes, and culminates with an exploration of its potential to contribute to learning (Cardoso, 2022). In this literature review, the handful of studies that have investigated IPAs' pedagogical potential in second language education will also be presented, following this chronology.

Given that IPAs are emergent worldware (i.e., technology not developed for educational purposes but repurposed for that use), very little is known about them and their pedagogical potential. Chronologically speaking, the first studies to explore IPA's pedagogical affordances and assess their suitability for second language teaching (stages 2 and 3) were Moussalli and Cardoso (2016), Dizon (2017) and Underwood (2017). Overall, these groundwork studies have found that the use of IPAs are perceived positively by users, particularly because of their interactive features. For example, the first known study on the pedagogical use of IPAs was conducted by Moussalli and Cardoso (2016; discussed thoroughly in chapter 2), a pilot study that investigated learners' perceptions toward the use of Echo as a pedagogical tool for English learning. The results showed that the adopted IPA was perceived by the participants as fun, easy to set-up and use; it created authentic contexts for human–machine interactions and constituted an effective tool to address learners' needs (e.g., personal pronunciation problems). Similarly, inspired by Moussalli and Cardoso (2016), Dizon's (2017) case study with four Japanese EFL learners examined Alexa's ability to understand second language utterances under two conditions: learner-generated commands and interactive storytelling. The results revealed that although the accuracy of Alexa in understanding second language utterances was moderate, the learners commented that the IPA provided them with indirect pronunciation feedback, which increased the opportunities to orally interact in the target language. In addition, the use of the IPA provided students with more access to conversational opportunities in the target language.

Concurrently with Dizon (2017), Underwood (2017) examined whether IPAs could be explored as a tool to encourage children to speak in the second language by developing appropriate AI-assisted task designs, and in addressing issues related to IPA-assisted classroom management and inaccurate voice-recognition. The author examined existing Artificial

Intelligence technologies (i.e., IPAs like Amazon Alexa; Siri on a class set of iPads; Google voice search on the teacher's mobile phone and iPads) in their ability to provide children with satisfactory answers to some of the common questions asked in language classrooms. The results revealed that the students found speaking to IPAs highly engaging, and they tended to speak English more frequently when using IPAs in group work. The findings also showed that the participants spontaneously reformulated, self-corrected, and persisted in speaking in English in their attempts to convince the IPAs to do what they wanted them to do.

A more recent study focusing on assessing IPA's pedagogical suitability (stage 3) was conducted by Wu et al. (2020). The authors contrasted the IPA user experience of non-native learners of English (L2) with native English speakers (L1) across two types of devices: smartphone and smart speakers. A within-participant design was used, where participants interacted with the IPA Google Assistant through both a smartphone (Moto G6 - Smartphone condition) and a Google Home Mini smart speaker (Smart speaker condition). According to the author, the results showed some differences in experience between L1 and L2 speakers when interacting with the IPA irrespective of the device type. For example, L2 speakers preferred using IPAs through smartphones because it provided them with support due to the provided visual feedback and the supplementary information displayed on-screen in response to the queries while L1 speakers preferred using Google Assistant, the smart speaker.

A more recent study on the use of IPAs to assess its effectiveness (stage 4) for second language listening comprehension and speaking development was conducted by Dizon (2020). The results highlighted the potential of IPAs to support foreign language development and showed that learners' perceptions were moderately favorable (enjoyed using the IPA and considered it to be a useful tool to study and practice English). In addition, Dizon's findings

showed some gains in speaking, especially on the participants' ability to produce comprehensible speech through the use of appropriate vocabulary, grammatical accuracy, and pronunciation, but not on listening comprehension.

Considering the small number of studies examining IPAs' pedagogical potential in providing learners with language input and allowing interaction with learners, my aim is to address this gap in the literature by extending the scope of the existing research and contribute with data and analyses that explore IPAs' pedagogical potential. Therefore, this dissertation contributes to the CALL literature by examining IPAs' pedagogical suitability, learner's perceptions towards them, and their potential to contribute to learning.

**The Current Dissertation**

The present PhD dissertation for a manuscript-based PhD dissertation adopts a commercial and widely available smart speaker, Amazon Echo, and its associated IPA and App, Alexa. I hypothesize that the adopted IPA could help improve second language learners' speaking and listening kills because it enables them to speak with it comfortably while listening to the IPA's output, thus encouraging oral production and aural perception in a self-directed manner (Johnson et al., 2014). It can also provide learners with visual feedback in the form of transcriptions or supplementary information displayed on-screen in response to their requests. I also hypothesize that the adopted IPA will be a valuable pedagogical tool in the second language classroom because of its ability to serve as a conversational partner, particularly if used as an extension of the language classroom (e.g., in combination with standard in-class teacher-facilitated practices). And finally, IPAs fulfil the criteria for CALL materials development set forth by Chapelle (2001), discussed earlier.

The overarching goal of this PhD dissertation is to examine a popular IPA (Amazon Echo

and its voice, Alexa) as a pedagogical tool in second language classrooms following the chronology for examining CALL technologies described in Cardoso (2022) by: exploring the tool (stage 2), examining its pedagogical suitability (stage 3), and assessing its pedagogical effectiveness (stage 4). Accordingly, the dissertation is divided into three main topics, each constituting a manuscript guided by a research question reflecting this chronological order: (1) perception of the tool from the learners' perspective, (2) appropriateness of tool in use (e.g., can the technology understand and be understood by its users?), and (3) potential to contribute to learning (i.e., learning gains in pronunciation):

1- Manuscript A (Chapter 2): What are language learners' perceptions of the use of IPAs as learning tools?
2- Manuscript B (Chapter 3): Can IPAs understand accented language learners, and can these learners understand IPAs?
3- Manuscript C (Chapter 4): Can IPAs help language learners improve their aural and oral skills (i.e., when learning about past -ed pronunciation)?

The first manuscript, exploring the tool (stage 2) and its suitability (stage 3), investigates the use of IPAs in general and examines users' perceptions of the technology as a language learning tool. It examines a number of variables such as the IPAs' ease of use, options for learner self-regulation (defined as learners' ability to understand and control their learning environment), learner motivation and, more importantly, opportunities for learner input (via listening) and output practice (via speaking). The second manuscript, assessing suitability (stage 3), examines the chosen IPA's ability to interact with different accented language learners of English. The focus is on exploring the IPA's ability to understand speech from different levels of L2 accentedness. It should be noted that the term understand is defined as IPA's ability to both

recognize a user's request and respond appropriately to the user. Finally, the third manuscript, assessing pedagogical effectiveness (stage 4), investigates whether the pedagogical use of IPAs can lead to improvements in learners' phonological awareness, perception and production of the allomorphy that characterizes regular past tense -ed marking in English (e.g., depending on the preceding phonological environment, suffix -ed can be pronounced as talk/t/, play/d/ and add/id/). In other words, this manuscript investigates whether IPAs can help learners notice, recognize, and discriminate the targeted -ed allomorphs, and produce them in controlled and guided settings.

## Chapter Two: Practicing English with a speaking robot: Learners' perceptions

### Introduction and Background

The use of technology to learn a second or foreign language (L2) has increased considerably over the last decade especially in Mobile-Assisted Language Learning (MALL). Briefly, MALL refers to language learning that is assisted or enhanced by the use of a handheld mobile device (Kukulska-Hulme, 2012). New technologies have accelerated changes in MALL environments by extending learning opportunities and reshaping learning styles where the lessons have become personalized, learner-centered, and collaborative (Kim & Kwon, 2012). This approach to learning is now known as Self Access Language Learning (SALL). SALL focuses on promoting learners' autonomy and independence so that learners manage their own learning (Gardner & Miller, 1999) outside the language classroom in a stress-free environment. This is particularly important in today's language classrooms, where students have limited time for oral practice and opportunities to communicate, especially in a stress-free environment (Neri et al., 2003). The limited classroom time learners have for oral practice (including listening/perception and production of speech) usually affects the quantity and quality of personalized feedback. Therefore, we believe the effective use of technology such as commercial (i.e., designed to be sold to the general public) and easily available Speaking Robots (SRs; e.g., Amazon's *Echo* and *Jibo*) in language learning might have the potential to promote acquisition by freeing up class time so that the instructor can focus on other important issues that require teacher intervention (e.g., teacher-facilitated interactions and individualized feedback), in a stress-free environment, which may then contribute to an increase in willingness to communicate (Gregersen & MacIntyre, 2014).

The main goal of this study is to explore the pedagogical use of SRs, specifically Amazon's *Echo*, as an effective and efficient tool to satisfy the needs of language learners, particularly when used as an extension of the classroom, in combination with standard in-class teacher-facilitated interactive practices. For this study, we adopted *Echo* (see Figure 2.1), a cylinder speaker that connects to its associated app, the cloud-based voice service *Alexa*, and instantly provides oral answers to any questions asked. E*cho* was selected because of its: (1) well-designed, tested, built-in Automatic Speech Recognition (ASR) software which transcribes speech and provides information based on an oral input; (2) low cost (approximately US$180); (3) popularity (e.g., see https://www.cnet.com/news/google-home-vs-amazon-echo); and, more importantly, (4) potential to satisfy the criteria proposed by Chapelle (2001) that researchers and teachers should consider when working with technology in second language learning. We summarize some of these criteria below, followed by related questions that should be asked to address them in the context of *Echo* (see the discussion section for a discussion of these criteria):

1. Reliability and learner fit: Is the use of the technology too easy or difficult for learners? How do learners' individual differences (e.g., their learning style, motivation, experience with technology) affect their performance?
2. Authenticity and generalizability: How does the use of the technology reflect technology use in the non-research environment? Can research results be generalized to other contexts?
3. Construct validity and operationalization of learning conditions: What theoretical constructs underlie use of the technology? How does use of the technology reflect those constructs?

4. Language learning potential and operationalization of learning conditions: What potential does use of the technology have for language learning? How is that potential realized in its use in research?
5. Interactiveness and meaning focus: How does the use of the technology engage learners' meaningful use of communicative abilities?
6. Positive impact: How do learners benefit or suffer from use of the technology?
7. Practicality: How easy is it to find, modify, and use the technology in a non-research context?

To summarize, we believe that the proposed pedagogical tool, *Echo*, is well equipped to provide second language learners with ample opportunities for oral production, and that the technology can be used to promote language learning by enhancing the quality and quantity of second language input, creating authentic contexts for communication to take place, and providing relevant and useful feedback (see Zhao, 2003 for a meta-analysis on the effectiveness of computer-assisted technologies in second language learning). Examples of technology that have the potential to be used effectively and efficiently in the language classroom include social networking (e.g., Facebook, Twitter), podcasting and, more recently, Text to Speech (TTS) and ASR (Liakin, Cardoso, & Liakina, 2015).

**TTS and ASR in the language classroom**

Text-to-Speech Synthesizers (TTS) are computer applications that convert text into speech. They are commonly used to assist users who cannot read or see, or to provide answers to voice-activated searches such as those found in browsers and GPS systems. In the case of voice-activated GPS (also considered a type of SR), for example, TTS is combined with Automatic Speech Recognition (ASR, also known as speech-to-text), a technology that allows users to

speak rather than type in information. To illustrate, consider a voice-activated GPS system. Based on the oral input received (e.g., a person's oral request for a location: "Where's Montreal?"), the built-in ASR system transforms that input into text, which is then used to look for answers on a text-based database (i.e., the web). After the information requested is located, the application's TTS capabilities output the answer in the form of speech. SRs' modus operandi resembles that of GPS systems and other voice-controlled personal assistants such as Apple Siri and Microsoft Cortana, whose goal is to recognize intelligible speech with accuracy and efficiency independently of the speaker's accent, background noise, and other variables.

There is little research that examines the use and potential of TTS and ASR, but existing research has found positive effects on language performance and oral self-assessment. The handful of studies that have investigated the effectiveness of TTS in second language acquisition of pronunciation suggest that this technology has the potential to enhance the acquisition of writing (Kirstein, 2006), vocabulary and reading (Proctor, Dalton & Grisham, 2007), and pronunciation (Cardoso, Collins & White, 2012; Handley & Hamel, 2005; Soler-Urzua, 2011).

Regarding the effectiveness of ASR, the existing studies have yielded positive results. Examples include Holland, Kaplan and Sabol (1999), who showed that the students' production of Arabic utterances improved as a result of an interactive computer program with a built-in ASR. Coniam (1999) evaluated the ability of the voice recognition software, *Dragon Naturally Speaking*, to recognize Cantonese-accented speech. Coniam (1999) showed that the software could be used to provide corrective feedback despite the fact that it was less effective in recognizing the speech of non-native speakers. Rogers, Dalby, and DeVane (1994) also looked at the effectiveness of minimal pairs drills using ASR. The authors used a word recognizer (not developed for language training) to determine whether feedback derived from a speech

recognition score could improve the intelligibility of second language speech. Their results showed that training with ASR was effective and that the improvement in intelligibility observed among ASR users was generalized to untrained words.

To assess the quality of ASR as a pedagogical tool, Derwing, Munro and Carbonaro (2000) evaluated the accuracy of *Dragon NaturallySpeaking Preferred*, a popular software application, in the context of high-proficiency English speakers. Their results reveal that the ASR was able to provide helpful negative feedback in a non-threatening context, although it was less accurate in recognizing non-native speech, similar to what was found in other studies (e.g., Coniam, 1999). These authors conclude that the ASR adopted in their study could not be considered a reliable tool for feedback on intelligibility, but they advocate its potential use in noticing activities to focus students' attention on their own production errors.

Unlike previous studies that used desktop-based ASR technologies, Liakin et al. (2015) adopted *Nuance Dragon Dictation*, a mobile ASR application, to investigate the L2 acquisition of certain phonological features in French. Their findings reveal that the use of a popular ASR-based dictation application yields a positive effect on the acquisition of the French vowel /y/ (as in "tu"). Their results also show that the participants enjoyed using the mobile ASR application to improve their L2 pronunciation and that it accommodated their learning styles (e.g., it promoted "anytime anywhere learning" and allowed participants to consistently and repeatedly test hypotheses about their own pronunciation). Similarly, Hsu (2015) reports in his study that visual and kinaesthetic EFL learners benefit from ASR-based pronunciation training because ASR accommodated their learning styles. He explains that these EFL learners who aimed to improve their pronunciation through self-regulated learning with ASR maximized their learning by selecting the features that matched their learning styles. Another example of the use of mobile

phones and learning styles is a study by Thornton and Houser (2005), who show that their learners using mobile technology preferred learning and having English vocabulary lessons on their mobile phones as opposed to desktop computers or traditional methods of content delivery.

The findings of the studies reviewed here suggest that the combination of ASR and TTS, when properly used, can be a valuable addition to face-to-face language teaching. Considering the potential of TTS- and ASR-based technologies for pronunciation teaching and learning, and the fact that there are no studies that have investigated the pedagogical potential of SRs in second language education, this study aims to address this gap in the literature by examining L2 learners' perceptions of one particular SR (Amazon *Echo*) as a pedagogical tool. As such, this study was guided by the following research question: What are English L2 learners' perceptions of using a speaking robot (Amazon *Echo*) as a pedagogical tool? To answer the question, our analysis included a quantitative analysis of survey results, using descriptive statistics, and a qualitative analysis of open-ended oral interviews with the participants, as will be described in the next section.

## Method

### Participants

This paper is a feasibility study that aims to explore learners' perceptions of the speaking robot *Echo* as an effective and efficient pedagogical tool. Therefore, the learner sample was small (N=7). All three female adult participants and four female adolescent participants were francophones and non-native speakers of English. The adult participants were students from French universities in Montreal while the adolescents were high-school students (N=2, age= 13 years) and elementary school students (N=2, age=11 years) from French schools in Montreal. The participants' proficiency level in English was intermediate, based on the researchers'

assessment and analysis of their responses to the language background questionnaire, developed for the purpose of this study.

**Amazon Echo and Alexa**

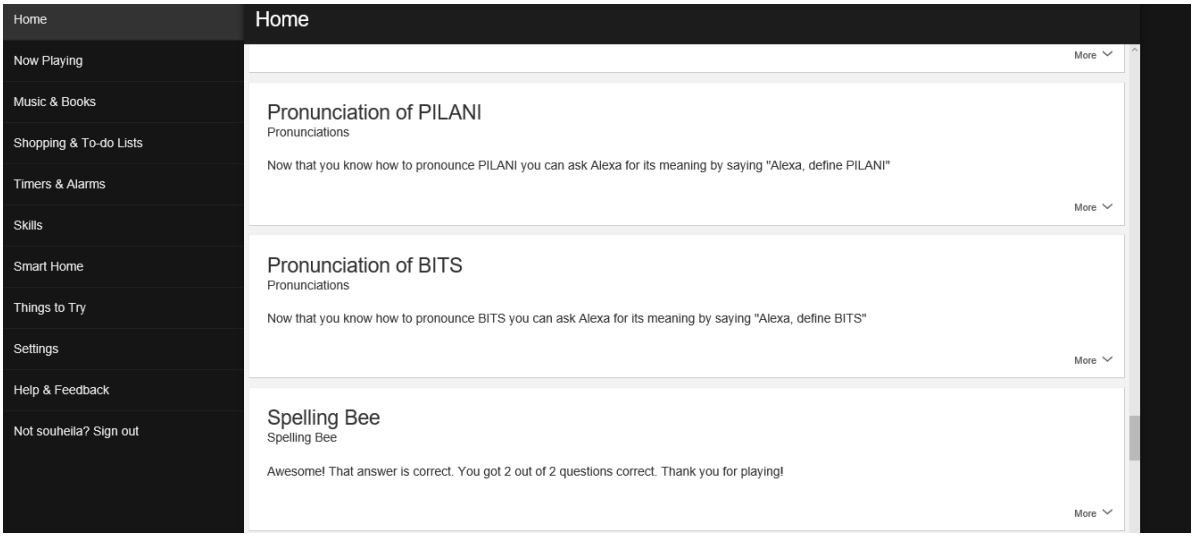
*Echo* is a tall cylinder (9.25-inch/23.5 cm) hands-free speaker, based on "beam-forming technology", a technology that allows speech signals to be heard (and consequently processed) by the device, regardless of the location of the interlocutor. Echo operates via its TTS- and ASR-based voice control system, *Alexa*. *Alexa* is a cloud-based voice service capable of voice interaction (as seen in Figure 2.2 below). When *Echo* detects the wake word, it lights up and streams audio to the cloud, where the *Alexa* Voice Service recognizes and responds to requests. In response to questions such as "How is the weather today in Montreal?", *Echo* is able to instantly provide information; it can answer questions, play music, read the news, check sports scores or the weather, etc. Another interesting feature is that *Echo* is speaker adaptive: The more a person uses *Echo*, the more its voice software *Alexa* adapts to the speech patterns, vocabulary, and personal preferences of the user.

**Figure 0.1**

*Amazon Alexa*

**Figure 0.2**

*Echo's Alexa App*




## Procedure

All participants initially filled out a consent form (parental consent was requested for the younger participants- Appendix A & B) and a background questionnaire (Appendix C), developed by the researchers, to report on their language learning experience and other personal information. They were then asked to interact with Echo for approximately 30 minutes using a set of questions and commands (n=26, related to general information and games-) and using their own questions (Appendix D). After their interaction with Echo, participants were given a 17-item survey (Appendix E) using a 5-point Likert scale (where 1 = strongly DISAGREE and 5 = strongly AGREE) to quantify their responses to a number of statements about their perceptions of the technology and its mode of operation (e.g., "Echo is able to understand me"). After the survey, a semi-structured interview was conducted by the researcher with each participant. The interview consisted of questions regarding the participants' performance and perceptions on the use of Echo and the Alexa App.

## Analysis

Means and standard deviations were calculated for each item of the survey used. As for the semi-structured interview, the researcher transcribed and analyzed the answers through topic-oriented cyclical data analysis (Watson-Gegeo, 1988). The interview questions were also categorized as related to one of the four topics of the research question, where responses can be relevant to more than one topic.

# Results

## Survey

For the survey, means and standard deviations were calculated for each item, as illustrated in table 1 below. Overall, participants enjoyed their experience using *Echo* as a pedagogical tool. To summarize the key points, the participants felt comfortable speaking in this type of computer-mediated communication (M=3.43), and considered it a great tool to learn languages (M=4). In addition, they found that Echo was able to comprehend their requests (M=3.57), was intelligible (M=4.14), and that their overall experience was enjoyable (M=4.57).

**Table 2. 1**

*Survey results: Learners' perceptions*

| Statements | Mean (M/5*) | SD |
|---|---|---|
| I felt more comfortable speaking English using Echo than I would in other types of classroom activities | 3.14 | 1.22 |
| I felt more comfortable speaking English while using Echo than I would in front of the teacher | 3.43 | .98 |
| I would like to use Echo to learn other languages | 3.86 | 1.22 |
| Echo is a great tool to learn a language | 4.00 | .82 |
| Echo is able to understand me | 3.57 | .54 |
| I was able to understand Echo | 4.14 | .38 |
| I enjoyed using Echo in this project | 4.57 | .79 |

* ranging from (1) strongly disagree to (5) strongly agree

## Interview

The transcriptions of the interviews conducted in English were analyzed by the researcher through topic-oriented cyclical data analysis (based on Watson-Gegeo, 1988), and categorized as relating to one of the four topics of the research question: ease of use, autonomy, opportunity for input and output practice, and motivation. This section ends with my report of the perceived weaknesses of the pedagogical use of SR and associated resources.

### *Ease of use*

Participants considered Echo user-friendly, enjoyable, helpful for language learning and fun: “It’s very clear it’s very simple, […] doesn’t have any complications”; “I enjoyed working with it, it’s very helpful”, “it’s the first time I talk with the machine, so I found it very um eh c’est amusant”; and “I was very motivated, I want to explore”.

### *Autonomy*

Participants found that Echo was helpful for acquiring certain pronunciation features and vocabulary on their own: “I can hear her to approve my English, so I can hear the way she talk and I learn from her the pronunciation”; “It increase the understanding of pronunciations and some vocabulary”.

### *Opportunity for input and output practice*

Participants expressed that the implicit feedback they received (e,g., “I did not understand” ) encouraged repetition: “I think it’s more encouraging, if you have to repeat, it’s like she don’t understand you, you can be better the next time”. Our findings also revealed that Echo was considered a helpful tool for the teacher to extend the reach of their classroom; for instance, by providing students with answers when their teachers were unable to help: “Yes,

because sometimes the teacher can not help you… for any reason, so… its can really helps", "Yes, I found it very helpful maybe in studies, in classrooms, to help teachers maybe".

***Motivation***

Most participants stated that their interactions with Echo motivated them to learn, to find out more about things that they were interested in, which can be exemplified in the following statements: "Yes, I was very motivated, I want to explore", "It pushed me".

***Weakness***

Despite the positive perceptions described above, the participants highlighted the weaknesses of the technology and their experience. For example, some participants questioned the use of machine-based interactions: "I prefer human interaction because I like… eh relation between student and teacher". They also reported problems with the speech recognizer, saying that it sometimes could not understand their requests: "It doesn't understand some words or some questions"; "Sometimes yes, sometimes no". Finally, most participants agreed that they did not like the fact that they had to use the wake word "Alexa" every time they wished to interact with *Echo*: "I don't like…before talking to her, you should talk the name Alexa".

## Discussion and Conclusion

This study investigated the pedagogical use of a SR (*Echo*) for L2 education and its potential to provide speaking/listening practice opportunities outside the classroom, and consequently improve learners' oral skills. The results of the study showed positive results for these analyzed variables: ease of use, options for self-regulation, motivation and opportunities for input and output practice.

With regard to ease of use, the results obtained here revealed that the participants enjoyed using Echo and found it very user-friendly. As previous CALL studies have suggested, this is an important feature for motivating the adoption of a new technology (e.g., Chapelle, 2001).

Participants felt that Echo helped them acquire certain pronunciation features and vocabulary, as explained in the following quotes: “Like even if I ask a question that I am not really sure… it can answer and it makes me understand more”, and “because sometimes I can ask myself what is the problem of speaking… yes it makes me thinking”. The forced output opportunity provided by Echo and the meaningful interaction with Echo might have prompted a gain in the participants’ performance with vocabulary and pronunciation (Gass, 1997). It is also possible that participants took advantage of the implicit feedback provided by Echo and enhanced their metacognitive knowledge about the features they were trying to learn by building on their existing knowledge (Pintrich, 2002).

Another interesting result relates to SR providing opportunities for input and output practice. As indicated earlier, participants expressed that Echo provided them with many chances for input and output practice, as expressed in the following quotation: “I will eh make a good sentence, maybe eh he, she doesn’t understand the way that I say the word, so I repeat in eh, in a different eh word”. However, the youngest participants felt that Echo did not provide them with many practice opportunities and at times they felt discouraged: “I try two times to ask again, but it doesn’t respond, so I stop”. This behaviour leads us to the possible conclusion that one of *Echo*’s main limitations is that it does not seem to adapt to younger voices well, as attested by these two participants’ comments. Further research is needed to investigate the effects of pitch and voice quality on Echo.

Finally, although participants agreed that, overall, using *Echo* was motivating, one adult participant expressed that using a machine-based interaction was less motivating. The results suggest that learning style, age group, and exposure are factors to consider when using technology in general, as attested in previous research (e.g., Czaja & Sharit, 1998, Hsu, 2015).

In conclusion, the results obtained in this study corroborate previous research in terms of ease of use (see Liakin et al., 2015 for similar results in a mobile ASR environment), assisting in providing pronunciation practice and its subsequent improvement, providing opportunities for input exposure and output practice (e.g., Derwing et al., 2000; Liakin et al., 2015; Neri, et al., 2006, 2008), and motivating students to learn on their own. We also acknowledge that participants' positive perceptions might be due to the novelty effect of the technology, which may have increased the overall interest and motivation of the participants (see Cardoso, 2010 for a discussion of the novelty effect). These results have implications for researchers', teachers', and students' use of this technology as a pedagogical tool (Chapelle, 2001). They can also serve as a starting point for a better understanding of a type of learning environment that is able to cater to different learning styles and that can extend the reach of the classroom and thus promote self-regulated learning. We hope that our study will pave the way for this under-studied area of CALL: the use of commercial (i.e., designed to be sold to the general public), easily available speaking robots as tools for L2 education.

The aim of this study is to investigate learners' perceptions of the pedagogical use of Amazon Echo and evaluate the potential it holds for language learning. The study closely investigated several variables such as Echo's ease of use, options for learner self-regulation or autonomy, learner motivation and, more importantly, opportunities for learner input and output practice. Given that there are only two studies that used IPAs for language learning and teaching,

and the anecdotal evidence that one of *Echo*'s main limitations is not adapting well to accented speech, the next study will explore the tool's potential further by examining Alexa's ability to understand and interact with different accented learners of English.

## Chapter Three: Intelligent personal assistants:

## Can they understand and be understood by accented L2 learners?

The second/foreign language (L2) classroom is notoriously known as an environment affected by time constraints (Collins & Muñoz, 2016; de Vries et al., 2015), which tend to result in a reduction of opportunities for students to practice and interact with others. To complicate the matter, the classroom is usually considered an anxiety-laden environment in which some students are anxious to communicate for fear of losing face (Neri et al., 2003) or because they deem their speaking abilities inadequate. As a result, students practice the target language less (including listening/perception and production of speech) and, consequently, the quantity and quality of the teacher- or peer-based feedback is reduced.

I believe that the pedagogical use of technologies such as Amazon's Echo and its voice controlled Intelligent Personal Assistant (IPA; other examples include Google Home and Apple's HomePod) can be valuable additions to "traditional" face-to-face L2 teaching because these devices have the potential to enhance in both quantity and quality the input that learners receive. In addition, they can motivate students to practice their newly-acquired oral skills and try their learning hypotheses—e.g., via IPAs built-in speech recognizer. IPAs can also be used to provide additional opportunities for communication outside the language classroom in a stress-free, meaningful and fun environment, anytime-anywhere, which may contribute to an increase in the learners' willingness to communicate using the target language (Gregersen & MacIntyre, 2014; Moussalli & Cardoso, 2016). Finally, and more importantly, the use of IPAs has the potential to promote learner autonomy, as the technology may expand learning opportunities to out-of-class contexts ("anytime anywhere" learning) and, at the same time, foster a personalized, learner centred, and collaborative L2 pedagogy (Kim & Kwon, 2012).

As a result, the scope and design of the current study assumes that IPA-based learning is self-regulated, which is informed by the self-access language learning (SALL) theoretical framework. SALL emphasizes the promotion of learners' autonomy and independence so that students can manage their own education (Gardner & Miller, 1999), especially outside the language classroom. SALL acts as a bridge between teacher-facilitated instruction, where language is learned and practiced during class time, and the 'out-of-class' world, where the target language is used as a means of communication (Cotterall & Reinders, 2001). In addition, due to the intrinsic nature of the human-machine interactions that characterize the pedagogy implemented in our study, it is also informed by interactionist approaches to L2 development, particularly regarding negotiations of meaning, in which the learner and the IPA engage in a continuing process of interactions to process and produce language (Long, 1996).

The goal of this feasibility study is to examine the capabilities of a popular smart speaker Echo and its voice-controlled IPA service Alexa for their ability to understand accented speech and be understood by English as a second language (ESL) learners. Specifically, it aims to address a limitation observed in a previous study, which showed that heavily accented English learners had difficulties understanding Echo's synthesized speech and being understood by its speech recognizer, Alexa.

**Intelligent Personal Assistants and Technology-Enhanced L2 Education**

In this study, I use the term IPA to refer to any of the popular voice-controlled services connected to smart speakers (e.g., Amazon Echo, Google Home, Apple's HomePod) that can perform various tasks and services by interacting with its users. According to Goksel-Canbek and Mutlu (2016), IPAs use a combination of three technologies: a conversational interface, personal context awareness, and service delegation to function and perform properly.

Conversational interfaces rely on automatic voice recognition tools (ASR) to interpret meaning. Personal context awareness enables IPAs to understand their users' language patterns and specific nuances. Finally, service delegation allows IPAs to integrate with other programs and apps. Both conversational interface and personal context awareness allow IPAs to communicate with users in human-like interactions (Johnson et al., 2014) because their built-in ASR system permits users to speak rather than type in information. In the context of voice-activated web searches, for example, based on a person's oral request for a location (e.g., 'Where is Montreal located?'), the built-in ASR system transforms that oral input into text, which is then used to look for answers on a text-based database (i.e., the web). After the information requested is located, the application's Text-To-Speech (TTS) capabilities output the answer in the form of speech. In other words, the IPA user will hear an oral reply to the original request. Using conversational interface and personal context awareness, IPAs combine user input, location awareness and ability to access information from a variety of online sources to perform the required tasks.

To our knowledge, there are only three studies that have investigated the pedagogical potential of IPAs in L2 education: a study conducted by Moussalli and Cardoso (2016), another by Dizon (2017) and a third by Underwood (2017). While the studies the Moussalli and Cardoso (2016) and Dizon (2017) examined learners' perception toward Echo and its associated voice Alexa, the latter also investigated the IPA's ability to understand L2 English speech. In particular, Dizon (2017) examined Alexa's ability to understand L2 utterances under two conditions: learner-generated commands and interactive storytelling. The results revealed that the accuracy of Alexa to understand L2 utterances was moderate, since the EFL-speaking participants struggled with being understood by Alexa through the learner-generated commands

more than through the interactive storytelling. The results also revealed that learner efficiency was hindered due to the lack of L1 support, preventing students from using their known language/s to aid the learning process. Overall, the pedagogical use of Alexa allowed students to receive indirect feedback and to increase opportunities to orally interact in the target L2.

The third study (Underwood, 2017) examined the existing Artificial Intelligence technologies (i.e., IPAs) in their abilities to provide children with satisfactory answers to some of the common questions asked in the language classrooms, and whether IPAs could be explored as a tool to encourage children to speak in the target language. The study also investigated ways of supporting children in developing appropriate AI-assisted task designs, and in addressing issues related to IPA-assisted classroom management and inaccurate voice-recognition. The results revealed that the students found speaking to IPAs highly engaging, and they tended to speak more English when using IPAs in group work. The results also showed that the participants spontaneously reformulated, self-corrected, and persisted with their use of English in their attempts to convince the IPAs to do what they wanted them to do

Thus, this study aims to address the gap in the literature by extending the scope of the existing research and by comparing the abilities of IPAs and English native speakers to understand accented L2 speech. Particularly, I would like to examine the following general question: How does Echo/Alexa and native-speaker judges (or raters) compare in their ability to understand accented speech?

I view IPAs as valuable tools because they can extend the reach of the classroom, motivate self-learning (particularly of speaking and listening skills and related vocabulary), encourage practice in a stress-free environment and, more importantly, they have the potential to promote L2 acquisition in general. In addition, the pedagogical use of IPAs may allow teachers

to relocate resources, particularly if the technology is used as an extension of the language classroom, in combination with standard in-class teacher-facilitated interactive practices. In the classroom, for example, teachers could focus on providing individualized feedback and teacher-facilitated instruction (e.g., for students struggling with the pronunciation of the vowel alternation/i/-/I/as in “beat” and “bit” in English, the teacher could ask them to interact with an IPA, using targeted questions, to find out about the spelling-to-sound rules involved). Finally, IPAs also meet the seven criteria put forth by Chapelle (2001) for selecting and adopting CALL tools:

1. Reliability and learner fit: Is the use of the technology too easy or difficult for learners? How do learners’ individual differences (e.g., their learning style, motivation, and experience with technology) affect their performance?
2. Authenticity and generalizability: How does the use of the technology reflect technology use in the non-research environment? Can research results be generalized to other contexts?
3. Construct validity and operationalization of learning conditions: What theoretical constructs underlie the use of technology? How does the use of technology reflect those constructs?
4. Language learning potential and operationalization of learning conditions: What potential does the use of technology have for language learning? How is that potential realized in its use in research?
5. Interactiveness and meaning focus: How does the use of technology engage learners’ meaningful use of communicative abilities?
6. Positive impact: How do learners benefit from use of the technology?

7. Practicality: How easy is it to find, modify, and use the technology in a non-research context?

I believe that the adopted smart speaker with its associated IPA, Alexa, meet these criteria. For example, it can be adapted to meet the learners' own interests and needs (learner fit), its use is authentic and can be generalized to other contexts (authenticity and generalizability), the pedagogy involved can be motivated by many theoretical constructs (e.g., Gardner & Miller's Self Access Language Learning, 1999; Long 's Interactionist Approach, 1996 ; Wong'sinput flood, 2005), it has great potential to promote learning via increased input and output practice (language learning potential), it fosters human-machine interactions that can be meaning focused (interactiveness and meaning focus), it can have a positive impact on learners' autonomy and general well-being (positive impact), and it is easy to use in out-of-the-classroom contexts (practicality). Some of these characteristics will become evident in the remainder of our discussion; others will be addressed in further research.

**The Target Smart Speaker and IPA: Amazon Echo and Its Voice, Alexa**

Amazon Echo is a cylinder speaker that connects to its associated app, the cloud-based voice service Alexa, and instantly provides oral answers to questions (also asked orally). It has a modus operandi that resembles that of other voice-controlled software such as Apple Siri and Microsoft Cortana. An interesting aspect of Echo is that its app, the Alexa App, provides transcriptions of the interaction between the user and Echo. It also allows users to enable different categories of skills that range from education, gaming, and lifestyle to business.
Echo was chosen for this study for a variety of reasons: It includes well-designed built-in ASR software that transcribes speech and provides information based on an oral input. It is relatively less costly than other similar devices (e.g., US$49 vs. US$350 for Apple's HomePod), it was

rated the best “smart speaker” in mid-2018 (Segan, 2018) and it was considered the most popular IPA in 2017 with a 70% share of the voice-controlled speaker market (Rao, 2017). More importantly, it satisfies the criteria proposed by Chapelle (2001), as discussed earlier; particularly, it is easy to use (e.g., it is voice activated, only requires an initial one-shot setup to function), has the ability to assist in pronunciation practice (e.g., it can serve as a model to teach vocalic alternations such as the/i/-/I/ alternation, described earlier), provides students with a large amount of input exposure and many opportunities for output practice, and it motivates students to learn on their own (Moussalli & Cardoso, 2016).

In a previous study by Moussalli and Cardoso (2016), Echo was used in a pilot study that investigated learners’ perceptions toward the use of Echo as an L2 pedagogical tool for ESL learning. The study looked at a number of variables that could affect learners’ perceptions: ease of use, options for self-regulation, motivation and, more importantly, opportunities for input and output practice. The results showed that the adopted system is fun, easy to set-up and use, creates authentic contexts for human–machine interactions, and constitutes an effective tool to address learners’ needs (e.g., personal pronunciation problems). The findings also revealed that Echo provided learners with speaking and listening practice in a stress-free environment, and offered helpful negative feedback, which stimulated learners to notice and identify problems or gaps in their production of the target language.

An important pattern observed in this previous study was that learners with low proficiency in English (high beginners), depending on their levels of accentedness, experienced difficulties understanding (listening) and being understood (speaking) by Alexa. This observation motivated the current study, in which I examine Alexa’s ability to recognize and understand accented speech to determine if it can effectively communicate with accented L2

speakers without causing communication breakdowns; the latter could potentially lead to app/device abandonment and potentially affect the overall acquisition process.

**IPA-Human Interaction and Accented Speech**

The past few years have seen a handful of studies on ASR's ability to recognize accented speech, especially because of our reliance on voice commands for searches on smartphones, a phenomenon commonly observed among the younger population (Enge, 2017). Despite this trend, the recognition accuracy of accented speech remains problematic because ASRs have difficulties processing non-native accented speech (Coniam, 1999; Derwing et al., 2000), which often results in communication breakdowns. In a more recent study, Enge (2017) surveyed 914 Americans by asking them whether their voice-activated personal assistants understood them. Fourteen percent of users claimed that they did poorly or very poorly understanding them, while 51% claimed that their personal assistant understood them well or very well. Whether this pattern is comparable to human-to-human communication will be discussed next.

**Human-human Interaction and Accented Speech**

There is empirical support for the claim that successful communication is not necessarily hampered by non-native accented speech (Derwing & Munro, 2009). Following Derwing and Munro (2009), I define accented speech as 'the way in which speech differs from the local variety of [that speech] and the impact of that difference on speakers and listeners' (p. 476). The concept of accentedness includes two sub-components: intelligibility and comprehensibility. Intelligibility is defined as 'the degree of a listener's actual comprehension of an utterance' (p. 478), while comprehensibility is 'the listener's perception of how easy or difficult it is to understand a given speech sample' (p. 479). In other words, accent is about difference,

comprehensibility is about the listener's perceived effort to understand, and intelligibility is the end result: how much the listener actually understands (p. 480).

There are many factors that may affect the comprehensibility and intelligibility of accented speech. Previous research on the relationship between these two constructs and accentedness show that they are overlapping yet distinct criteria. This means that not every highly accented speaker has low intelligibility or comprehensibility. In other words, it is possible for heavily accented speakers to be completely intelligible, but if a speaker is unintelligible, then for sure that speaker is rated as being heavily accented (Derwing & Munro, 2009). Derwing and Munro (1997) also show that accented ratings are usually harsher than intelligibility ratings; they explain that some highly salient features of accented speech, which are fully intelligible, are rated as difficult to understand because they need additional effort or processing time from their listeners. The authors conclude that accentedness does not necessarily impede comprehensibility and intelligibility to the degree that communication breakdowns would occur (Derwing & Munro, 2009).

Nevertheless, communication breakdown might occur in interactions with accented speakers. Previous research has shown that learners tend to use different negotiation strategies to overcome communication breakdowns when they occur. For example, learners might modify their speech in terms of form and meaning, repeat utterances, elaborate on utterances, adjust the utterance's syntax, change the vocabulary, etc. To resolve communication breakdowns, they may also coin new words (e.g., based on their L1), self-correct, use their first languages or L2 translations, or even change the topic altogether (Canale & Swain, 1980; Tarone, 1980). In addition to the verbal strategies that learners use as part of their negotiation of meaning to resolve communication breakdowns, non-verbal communication is also considered another

strategy used to achieve similar results. Non-verbal communication is 'the process of one person stimulating meaning in the mind of another person or persons by means of non-linguistic cues, e.g., facial expressions, gestures etc.' (Negi, 2009, p. 101). Research in non-verbal communication shows that effective communication is usually handicapped without the proper use of non-verbal behaviors (Canale & Swain, 1980), which help interlocutors enforce assumptions and emphasize the explicit meaning of a message (Ka łuska, 2013). According to Ozkan and Morency (2013), learners often influence each other during face-to-face communication through their verbal and nonverbal behaviors (also known as backchannel feedbacks) such as 'hmm or uh-huh' or head nods and facial expressions. These cues do not interrupt the conversation; instead, they allow the interlocutors to either continue their conversation or provide additional explanation based on the feedback received from their listeners (Ozkan & Morency, 2013).

Previous research that has looked at the different negotiation strategies used by learners have either looked at discourse functions (i.e., by counting the different types and quantity of negotiation), or followed the discourse structure proposed by Varonis and Gass (1985), which I adopted in this article to analyze the negotiation moves between learners and the target IPA. Varonis and Gass (1985) explain their proposal as containing four components: (1) Trigger: The initial word or utterance which initiates the misunderstanding; (2) Indicator: The Listener's signal that something was misunderstood; (3) Response: The Speaker's response to the Listener's signal; and (4) Reaction to the response: The listener's reaction to the Speaker's response. According to the authors, the horizontal flow of conversation is interrupted when the indicator (I) occurs following the trigger (T), after which negotiation starts until successful

understanding is successfully achieved, or not. After this point, conversation goes back to the main line of discourse.

**Human-machine Interactions**

In human–machine interactions, negotiation strategies can take place differently and, as examined in many human-machine studies, they trigger the same patterns and create the same favorable conditions as those in L2 human–human interactions (Smith, 2003). For example, a study on vocabulary acquisition using a synchronous chat by Blake (2000) revealed that negotiation was mainly due to lexical confusions. To resolve these confusions, which caused communication breakdowns, learners used explicit corrections, direct questions and requests, and extended probing. Relatively similar findings were observed in a study by Fernandez-Garcia and Martinez Arbelaiz (2002), who investigated written negotiation in a chat program. Their study also revealed that learners negotiated mostly over lexis and, to resolve their communication breakdowns, they resorted to strategies such as translations, direct questions, explicit statements of non-understanding, and echo (the repetition of the unknown item, followed by a turn with a question Regarding communication breakdowns in machine-learner oral interactions, a study by Fernandez-Garcia and Martinez Arbelaiz (2003) examined both oral and written texts in chats. Their findings revealed that non-native speakers negotiated significantly more in the oral than in the written mode, and in the former, they used strategies such as explicit statement of non-understanding, inappropriate or incorrect response, no verbal response, and rephrasing with rising intonation. In written texts, on the other hand, the participants used explicit statements of nonunderstanding, echo questions, or they appealed for assistance.

At least two studies on machine-learner oral interactions indicate that negotiation moves, as a results of communication breakdowns, were mainly due to pronunciation issues (Jepson,

2005; Lee, 2009). In these negotiations, learners used the same strategies observed earlier for this type of interaction: explicit statement of non-understanding, self-repetition, clarification requests, comprehension checks, explicit correction, confirmation checks, elaboration requests, inappropriate or incorrect response (indicating non-understanding), nonverbal response, elaboration, rephrasing, paraphrasing, surprise reaction, and self-repair.

A study by Yanguas (2010) compared learners in face-to-face (FTF) interactions with two types of video vs. audio-only computer-mediated communication (CMC) applications while completing a jigsaw task with unknown vocabulary. The results revealed that the negotiation instances for oral CMC were higher than those found in FTF and video-based interactions because oral CMC forces learners to make use of linguistic resources—e.g., in an audio-only environment, learners cannot resort to gestures and other non-verbal strategies to communicate. Yanguas (2010) also found out that negotiation was mainly over lexis because of the nature of the tasks—to complete a jigsaw task.

In sum, the negotiation strategies used in many human–machine interactions are similar to those found in human–human interactions. However, in human–human interactions, supra-segmental features (like intonation) and paralinguistic features (e.g., gestures, facial expressions, head-eye movements) are integral parts of successful communication, as they can help clarify the source of the misunderstanding. Oral human-machine interactions, on the other hand, are considered helpful in promoting oral production and conversation skills, considering that audio and video-based CMC (and possibly IPAs) resemble in many aspects the features found in human-human interaction.

## The Current Study

To discover the types of communication breakdowns that characterize machine-human in IPA-based interactions, and to address a limitation observed in previous research by the Moussalli and Cardoso (2016), namely that accented English learners experienced difficulties understanding and being understood by the IPA, the current study investigates Alexa's ability to recognize and understand accented L2 speech of different levels of accentedness. The study also aims to determine whether Alexa, the IPA adopted, can communicate effectively with accented speakers, without resulting in a communication breakdown. I thus examine how Alexa responds to L2 accented participants' speech in terms of accuracy in response, the number and types of communication breakdowns observed, as well as the effect of accented speech on their performance; i.e., how participants behave to solve the interaction problems that they experience with the IPA.

The study was guided by the following research questions:

(1) Can Alexa understand accented speech of English learners from different first languages?

(2) Can the same L2 learners understand the synthesized (and consequently also "accented") speech of Alexa?

(3) How do Alexa and native-speaker judges (raters) compare in their ability to understand accented speech?

(4) When Alexa-learner communication fails, what strategies do the learners use to solve the communication breakdown?

## Method

In this study, participants were asked to interact with the IPA using a series of questionnaires and then, they were asked to rate Alexa using a survey and transcribe a sample speech from the IPA. All instruments and procedure will be explained thoroughly in the forthcoming sections.

### Participants

Eleven English L2 learners (5 males and 6 females; age range: 19–30) from different language backgrounds (French, Cantonese, Mandarin, Arabic, Hindi, Tulu, Marathi-Gujarati) were recruited from a university in Montreal, Canada—their participation was on a voluntary basis. Only participants who met our two main criteria (level of accentedness and English language proficiency) were selected for the study, regardless of whether they favored the use of technology or not for language learning purposes. This information was gathered and verified based on the language background questionnaire that the participants filled out at the beginning of the study, and the researchers' assessment of their oral abilities when they interacted with Echo. Hence, the participants' proficiency level in English varied from low intermediate to advanced.

### Instruments

#### *Questionnaire*

At the beginning of the experiment, the participants were asked to fill out a consent form (Appendix F) and a language background questionnaire to gather demographic data and other relevant information such as their previous language learning experience. The questionnaire consisted of 11 questions in English about the participants' first languages, their language learning history, and their knowledge and use of technology (Appendix C).

### *Surveys*

The participants were asked to complete a 17-item survey (Appendix E) using a 5-point Likert scale (ranging from (1) strongly disagree to (5) strongly agree), which aimed to quantify their responses to several statements about their perceptions of their pedagogical experience using the target IPA (e.g., 'Alexa can understand me'). The survey was adapted from Liakin et al., (2015) and, as customary in survey studies, the methodology for designing and validating its items followed standard research methods protocols (e.g.,, it was checked for internal consistency using Cronbach's Alpha., and was later revised based on pilot-testing and users' feedback). In addition, for this study, all questions were pilot-tested by a small group of participants and the two most relevant questions (i.e., 'Alexa can understand me' and 'I can understand Alexa') were consistently deemed clear, easy to understand and, more importantly, directly related to answering whether IPAs can understand (#1) and be understood (#2) by accented L2 learners. Due to the scope of the current study, which aimed to answer this question via a triangulation of methods (including surveys, judges' rating on two pronunciation measures, and interviews), I only report and discuss the results obtained for these two relevant items.

### *Participants' ratings and transcriptions of the IPA's speech*

To measure comprehensibility and accentedness, as is customary in the L2 pronunciation literature, the participants were asked to rate on paper Alexa's speech globally on a scale from (1) difficult to understand or unnatural to (5) easy to understand or natural, respectively. To measure intelligibility, participants were asked to transcribe (like in a dictation) Alexa's reply to the question 'Alexa, are you in love?' (i.e., "I'm totally cool with being single. Besides, it's sort of hard finding someone who's kind, funny, artificially intelligent, and doesn't mind the cloud commute") (Appendix G); for an overview of these techniques for measuring comprehensibility

and intelligibility, see Derwing and Munro (1997). Note that the target response contains both frequent and rare words (27 unrepeated unique words, 81.5% of the words were in the K1 band, and 11.1% of the words were in the K2 band, with the rest being words from the Academic Word List or off-list words, with a type-token ratio of 1.0), simple -and complex syntax, thus giving participants ample opportunities to demonstrate how much they actually understood (or not) Echo's response.

***Interview***

A semi-structured oral interview was conducted by the researcher with each participant for an average of 30 minutes after the data collection phase (see forthcoming description). Questions included: 'How did you like your experience with Alexa?'; 'Did you understand Alexa? Was it too fast or slow?'; 'Did Alexa understand you? If no, how many times did you have to repeat until it understood you? The purpose of the interview was to obtain more in-depth insights about the participants' experience with the IPA.

***Judges' ratings and transcriptions of participants' speech***

Two English native speaker judges with limited exposure to the participants' first languages rated 15 randomly selected speech samples that represent different types of Echo-learner interactions, using a 5-point Likert scale on comprehensibility and accentedness (see Appendix H). These speech samples of accented L2 speech were extracted from the participants' corpus and they consisted of sentences/questions (e.g., 'Alexa, how far is the moon?'; 'Alexa, repeat after me: …") with an average of 4.2 words in length, mostly using simple present and future constructions. The rationale behind the inclusion of these judges' rating is to compare their assessment of the participants' comprehensibility and accentedness with that of Alexa. In other words, to further triangulate the data analysis, I will compare Alexa's performance with

that of human whenever paralinguistic features such as gestures and other visual cues are not present. Finally, to determine differences in intelligibility between the native-speaker judges and Alexa, transcriptions from the judges and those provided by the Alexa App will be compared to calculate percentages of transcription accuracy.

**Treatment**

During the data collection stage, the participants were asked to interact with Echo for approximately 45 minutes, during which they were video recorded. They were provided with a pre-established set of questions created by the researchers, but they were also encouraged to create and ask their own questions after they familiarized with Echo and its modus operandi. The pre-established questions (n=30) included language-related queries such as defining a term (e.g., 'Alexa, what is the definition of convoluted?') and/or spelling words (e.g., 'Alexa, how do you spell bits/beats?'), translating into other languages (e.g., Alexa, how do you say love in French?'), other general learning questions (e.g., 'Alexa, how far is the moon?', 'Alexa, what is the square root of 64?'), and games (e.g., 'Alexa, play rock paper scissors'). Appendix D provides a list of the target questions used in the study.

While interacting with Alexa, the participants were provided with two types of feedback: aural and written feedback. While aural (implicit) feedback was provided by the participants' listening to and interpreting Alexa's answers (e.g., if Alexa did not respond or provided an incorrect answer, that indicated that the interaction was not successful), the written feedback was provided through the transcription of the interaction via the Alexa App, which was made available to the participants after they completed the assigned tasks but before engaging in an oral interview.

After the participants interacted with Alexa and completed all their tasks, the researcher conducted a semi-structured oral interview with each of them for an average of 30 minutes. These qualitative data were transcribed and analyzed by the researcher according to the coding methods proposed by Saldaña (2009), which recommends that the analysis should move through different coding cycles: from creating codes to general categories in the first cycle, to narrowing it down to more specific themes and concepts in the second cycle.

**Analysis**

To measure how much Alexa can understand the participants' accented speech (research question 1), a combination of responses to a survey question, an analysis of Alexa-produced transcripts (to determine levels of intelligibility), and an interview question in which learners were asked to voice their perceived difficulties in communicating with the IPA were employed. The second research question (Can L2 learners understand the synthesized speech of Alexa?) was addressed via another triangulation of methods that consisted of a survey question, the participants' ratings of Alexa's voice on comprehensibility and accentedness, their transcriptions of samples of synthesized speech (again, to determine levels of intelligibility), and their stated perceptions of the IPA's voice. For the third research question (How do Alexa and native-speaker judges compare in their ability to understand accented speech?), comparisons between Alexa's and the judges' levels of understanding (intelligibility) was established by contrasting the transcriptions produced by Alexa (via its app) and those submitted by the judges. Finally, the last research question (What strategies do the learners use to solve the communication breakdown?) was examined via a computation of all communication breakdowns (e.g., if Alexa did not respond or provided an incorrect answer, that indicated that the interaction was not successful) and their classification according to Varonis and Gass' (1985) model for analyzing

the phenomenon. Within this approach, interactions was coded as Triggers (initial word or utterance which initiates the misunderstanding), Indicators (the listener's signal that something was misunderstood), Responses (the speakers' response to the listener's signal), and Reactions to Response (the listener's reaction to the Speaker's response).

Due to the nature and scope of this feasibility study, descriptive statistics was used for the instruments that include quantitative measures; these include survey questions and participants' and judges' ratings for comprehensibility and accentedness, for which means and standard deviations were also computed. To assess the inter-rater reliability of the judges' ratings of learner speech, Cohen's kappa coefficient statistics was used. For results regarding intelligibility, percentages of transcription accuracy were calculated from both Alexa and the judges by subtracting the number of words incorrectly transcribed (missing words, or incorrect spelling of the words) from the total number of words in the sample. In addition, I calculated the percentage of Word Error Rate (WER) of a sample from Alexa's transcriptions, which constitutes a way to measure the adopted IPA's performance as a speech recognizer. WER was calculated by adding the number of words deleted (D), substituted (S) or incorrectly inserted (I) by the ASR, dividing that sum by the total number of words in the sample (N), and finally multiplying the results by 100 for a percentage value (WER=S+D+I/N*100).

The qualitative data were transcribed and analyzed according to the coding methods proposed by Saldaña (2009), which recommend that the analysis should move through different coding cycles: from creating codes to general categories in the first cycle, to narrowing it down to more specific themes and concepts in the second cycle. Briefly, the participants' answers were initially compiled by the researcher into two main categories about the strengths and weaknesses of the pedagogical experience. I then divided the general topics into subcomponents that more

directly reflected the goals of the study (e.g., within the general "weaknesses" code, whether Alexa was perceived as hard to understand). This coding method helped us organize the participants' intended meaning for analysis.

**Results**

To address our first research question: Is Alexa able to understand the accented speech of English learners from different first languages? We first analysed the two survey items that focus on the participants' ability to understand and be understood by Alexa, as illustrated in Table 3.1, participants found that Echo was able to understand them relatively easily (M=3.55/5, SD=1.63).

**Table 3.1**

*Survey: Can Alexa understand and be understood by accented L2 speech?*

| Statements | Mean/5* | SD |
|---|---|---|
| *Alexa* can understand me. | 3.55 | 0.93 |
| I can understand *Alexa*. | 4.18 | 0.75 |

*1=strongly disagree to 5=strongly agree

These findings were corroborated by an analysis of Alexa-produced transcripts (which served to determine levels of intelligibility), in which we observed that the IPA's error rate was 23%. This result was calculated from the Word Error Rate (WER) of a sample from Alexa's transcriptions (7 participants), which is a way to measure this IPA's ASR performance. Finally, similar patterns were observed in the semi-structured interview, in which the participants expressed that Alexa understood them well: e.g., 'considering my accent, yes, she [Alexa, a female voice] understands right', 'Yes, it almost answered everything'. Similar comments were observed in the interview data of all participants.

Regarding the second research question (Can L2 learners understand the synthesized speech of Alexa?), the survey results shown in Table 3.1 indicate that the participants believed that they could understand Echo without difficulties (M=4.18/5, SD=0.75). As expected, similar

results were also obtained based on the participants' ratings of Alexa's speech, reaching M=4.55 for comprehensibility and M=4.45 for accentedness. For intelligibility, the rate of correctly transcribed words was 75% (only 25% of the words were incorrectly transcribed), as shown in Table 3.2. Accordingly, these findings were supported by the participants' comments stating that they had no or little problem understanding Alexa (most of their comments were about speed or the robotic nature of the speech): e.g., "Yes, it's good, not fast" 'Yes, Nice voice, sometimes a little bit robotic"; "Yes, it's good, not fast, not slow".

**Table 3.2**

*Comprehensibility, Accentedness and Intelligibility: L2 learners' ratings of Alexa*

| | L2 Learners |
|---|---|
| Comprehensibility (M /5) | M= 4.55 (0.52) |
| Accentedness (M /5) | M= 4.45 (0.52) |
| Intelligibility (%) | 75% |

*Note.* 1 = difficult to understand or heavily accented; 5 = easy to understand or not accented

To answer the third question (how do Alexa and native-speaker judges compare in their ability to understand accented speech?), we compared the transcriptions produced by Alexa (via its app) and those submitted by the judges. To ensure homogeneity among the two judges, we used Cohen's kappa coefficient statistics to determine inter-rater reliability for accentedness (ICC = 0.588), comprehensibility (ICC = 0.576) and intelligibility (transcriptions): exact word match of 84.6% / Cohen's kappa $\kappa$= 0.567; these results suggest a moderate level of reliability. Means and standard deviations for comprehensibility and accentedness for the 15 items rated by the two judges was found to be M=3.7/5 (SD=0.97) and M=2.46 (SD=0.84) respectively, as shown in Table 3.3. For intelligibility, the results for transcription accuracy indicate that Alexa and the raters understand L2 accented speech with relatively similar accuracy, above 80%.

**Table 3.3**

*Raters vs. Alexa: Comprehensibility, Accentedness and Intelligibility of L2 Learners*

| | Raters | Alexa |
|---|---|---|
| Comprehensibility (M /5) | M= 3.7 (0.97) | N/A |
| Accentedness (M /5) | M=2.46 (0.84) | N/A |
| Intelligibility (%) | 95% (N/A) | 83% |

To answer the fourth and last research question (i.e., when IPA-learner communicative interactions fail, what strategies do the learners use to solve the communication breakdowns?), we transcribed and coded the IPA-participants interactions based on the model proposed by Varonis and Gass (1985), as described earlier. We found that, out of the 1,000 interactions that occurred between the IPA and the participants, the total number of communication breakdown was 177. The results revealed that the communication breakdowns observed were mainly caused by pronunciation issues (94/177 = 53.11%, indicated by *), as summarized in Table 3.4.

**Table 3.4**

*Types of communication breakdown (CB)*

| Type of CB | Examples | Total (177) |
|---|---|---|
| Incorrect sentence structure | From Montreal and Quebec, what is the distance between? | 33 |
| *Pronunciation error: segments | How many cups in a liter (lighter: [lajtər])? | 40 |
| *Hesitations, stammering | could you… help me… with pronouncing b.i.t…s? | 37 |
| *Very fast speech/slow speech | N/A | 17 |
| Phrases not requiring a response | Wow, that's great! | 11 |
| Grammatically complex questions | I'm thinking what to have for lunch. Suggest something which is Mexican cuisine. | 11 |
| Atypical demands | Can you shout for me?! | 28 |

When a communication breakdown occurred, participants behaved differently in terms of solving the problems after they received the indicator from Alexa in the form of implicit feedback (e.g., via a comment/question, silence, or an incorrect response). In those cases, participants tended to adjust the questions' syntax and/or change the vocabulary and pronunciation. The results in Table 3.5 show that the most prominent interactional moves (strategies henceforth, as commonly referred to in the field) used by the participants were: repeating the problematic question (n=77/170; 45.3-%), re-phrasing it (n=54/170; 31.76%), and finally abandoning it (n=39/170; 22.94%) to resolve the communication breakdown.

**Table 3.5**

*Participants: Resolving interaction problems*

| Type of Behavior | Total (n=170) | % |
|---|---|---|
| Repetition | 77 | 45.30 |
| Rephrased | 54 | 31.76 |
| Abandoned | 39 | 22.94 |

The following example illustrates one of the commonly used strategies the participants employed to cope with communication breakdowns (rephrasing):

| | | |
|---|---|---|
| (Trigger) | Participant: | Alexa, where is located… Niagara Falls? |
| (Indicator) | *Echo*: | Hm, I can't find the answer to the question I heard. |
| (Response) | Participant: | Alexa, where is Niagara Falls located? |
| (Reaction) | *Echo*: | Niagara Falls, New York, is a waterfall in Niagara county... |

## Discussion and concluding remarks

This study investigated a popular IPA's ability to recognize and understand accented L2 speech and, consequently, to determine whether it can communicate with heavily accented

learners without resulting in a communication breakdown, as has been reported in research on human-to-human interactions (e.g., Derwing & Munro, 2009). The findings related to the first and second research questions suggest that the adopted smart speaker and its IPA, Alexa, can easily understand accented speech from English learners of different first languages. Although the accentedness rating from the judges was 2.8/5, slightly above the neutral 2.5 mark, this did not seem to affect the IPA's ability to understand the participants' accented speech. A possible explanation for this observed pattern may be that, in non-native speech, accentedness ratings are usually harsher than intelligibility ratings (see Munro, Derwing & Morton, 2006 for similar observations), and features that contribute to a perceived non-native speakers' accent do not always influence comprehension (Ortega-Llebaria, 1997).

Despite the fact that the observed word error rate for speech recognition technology is assumed to be 5% or less (Protalinski, 2017), the findings observed in this study suggest that Alexa is able to understand accented speech even at a 20% word error rate. From a pedagogical point of view, having a high error rate is a positive outcome; it suggests that Alexa can accommodate to learners' pronunciation and/or lexis without relying on strategies that do not require input processing, such as guessing from linguistic context or other non-verbal cues, as humans do when interacting with each other, or as the judges did in this study. For example, even though Alexa did not understand the question 'How many cups in a liter (pronounced [lajtər])?', our judges had no problems understating it, as indicated by their accurate transcriptions. This suggests that, while our judges relied on the context and their own language experiences to understand the target accented speech, the IPA relied mostly on the accuracy of the speech. Consequently, when Alexa detects these pronunciation/lexis issues and provides learners with feedback in the form of a comment, question, silence, or an incorrect response,

learners may realize that their utterance is inaccurate, and they are thus encouraged to improve their pronunciation in order to communicate. This encourages repetition with improvement sometimes, an important attribute that is claimed to contribute to L2 learning (e.g., see Celce-Murcia, Brinton, Goodwin, & Griner, 2010 for pronunciation, and Horst, 2013 for vocabulary). This type of implicit feedback stimulates learners to notice and identify problems or gaps in their production of the target language.

As for the third research question, our results indicate that learners used a variety of strategies to mitigate the communication breakdown they experienced with Alexa. The most prominent strategies consisted of repeating their questions, rephrasing them, and abandoning them, in that order. One explanation for this persistent pattern to have their utterances understood could be that the machine-human interactions via the IPA lowered the participants' affective filter because the device can never become impatient, bored or irritated. In these interactions, when Alexa did not understand their questions, the participants rephrased or repeated their questions, without the fear of losing face – machines are not judgmental. They made use of either the implicit feedback provided by the IPA (e.g., via a comment/question, no response, or an incorrect response) or the explicit feedback they received (via the app transcriptions) in order to produce a more coherent output. After a few trials and errors, participants abandoned the question, as a last resort. In this scenario, learners were likely stimulated to talk more in the target language, as they might have felt more relaxed (similar to what was observed with ASR-assisted learning, reported in Van Doremalen, Boves, Colpaert, Cucchiarini, & Strik, 2016), and less pressured to maintain successful communication (Fernandez-Garcia & Martinez Arbelaiz, 2002).

In sum, our results indicate that the adopted IPA adapts well with accented speech. In addition, it exposes learners to oral input that is abundant and of good quality, and provides them with ample opportunities for practice (both input/listening and output/speaking) through human-machine interactions. As observed in previous studies involving the pedagogical use of ASR (e.g., Derwing et al, 2000; Liakin et al., 2015; Neri et al., 2003; Van Doremalen et al., 2016), these affordances may motivate learners to explore the target language and consequently monitor their self-regulated language learning process.

## Limitations and Future Directions

While we recognize that this is a feasibility study and, accordingly, limited in scope in some respects, we would like to acknowledge some of its limitations and suggest directions for further research. The most important limitations include the small number of participants and judges, a lack of control for the participants' first languages, and their levels of accentedness in the target language (i.e., whether they would be considered as heavily accented or not). In future research, we would like to examine the IPAs' ability to understand accented speech and be understood by the same speakers, using a wider variety of first and target languages. Another limitation is that our study focused on one single IPA, Amazon Echo. To compare the affordances of IPAs in general and their individual limitations, and potentially suggest the ones that are more suitable for L2 pedagogy, we recommend that future studies consider a larger selection of IPAs. We should also examine the interaction of groups of L2 learners with an IPA and examine the strategies used while they collaborate with peers and interact with the device. Another aspect worthy of research is to investigate learners' and teachers' perceptions towards the use of IPAs for language learning purposes. Finally, an important area to be explored is whether the extended use of IPAs may contribute to the learning of specific language features

such as pronunciation, vocabulary, and grammar. Although previous studies on ASR have shown positive results on the acquisition of vocabulary (e.g., Thornton & Houser, 2005) and pronunciation (Liakin et al., 2015), it would be important to verify whether the pedagogical use of IPAs would contribute to an improvement in L2 fluency, prosody (e.g., intonation for questions, rhythm) and segments (e.g., the vocalic /i/-/I/ contrast, as in “beet” and ‘bit” respectively, based on explicit form-focused learning with an IPA).

This feasibility study has shown that intelligent personal assistants such as Amazon Echo and its synthesized voice, Alexa, can understand accented L2 speech and be understood by the same accented L2 learners, with a performance relatively similar to that of two human judges. Based on our findings, we conclude that IPAs may be used as valuable teacher assistants that can interact with students and motivate them to learn on their own, at their own pace, with great potential to extend the reach of the L2 classroom and consequently encourage practice and anytime-anywhere learning.

## Chapter Four: Intelligent Personal Assistants in L2 Education: Focus on English Past -ed

The progress and advent of new technologies have had a positive effect on language teaching and learning in general (Friggard 2002; Mansor 2007; Timucin 2006), as its use brings variety into the classroom, provides learners with increased autonomy and opportunities to regulate their own learning, and develop specific language skills while offering easy access to information outside the traditional learning environment (Braul, 2006). Technology also allows teachers to enhance their classroom practices and boost their teaching methods, where more emphasis is placed on engaging interaction and meaningful communication, especially for the teaching of grammar (Lam & Lawrence, 2002). According to Godwin-Jones (2009), today's expectation for technology-enhanced pedagogy is that the process is not an isolated, stand-alone activity, but rather integrated into a communication-centered, networked language learning environment. Heift and Vyatkina (2017) also add that grammar teaching with technology should focus on how learners use and interact with it and their peers as they complete tasks, allowing learners to explore authentic language, promote independent discovery and learner autonomy. Godwin-Jones (2009) further explains that technology for second language (L2) teaching and learning should focus on learners' attention to forms and structures and the accompanied grammar exercises need to be integrated, intelligent, and innovative.

There are a number of studies that have looked at the efficacy of using technology for grammar acquisition, focusing mainly on improvements in syntactic accuracy or complexity. For example, Shaalan (2005) created an Intelligent Language Tutoring System for learners of Arabic, which was designed to provide written feedback on grammatical errors using Natural Language Processing. The studies that investigated the use of technology in the teaching of grammar have showed positive results inasmuch as they helped improve learners' grammatical accuracy of the

target features (for further detail see Al-Jarf 2005; Baleghizadeh & Oladrostam, 2010; Fiori, 2005; Mohamad 2009; Nagata 1996; Nutta 1998; Pellettieri 2000; Pirasteh 2014; Salaberry 2000; Sotillo 2000; and Wang & Smith 2013).

Nevertheless, none of the abovementioned studies have examined the influence of technology on the *pronunciation* of grammatical forms (morphophonemics), as they focused on grammatical accuracy, as previously stated. According to Celce-Murcia et al. (2010), many English grammatical forms are affected by their phonological environments. Consider past -ed marking, for example, a phenomenon that is dependent on the preceding phonological environment: -ed is pronounced /t/ when preceded by voiceless segments, /id/ if the preceding consonant is /d/ or /t/, and /d/ elsewhere. For example, Cardoso (2018) explored the pedagogical use of a text-to-speech (TTS) synthesizer in helping English as foreign language students learn the pronunciations associated with regular past tense marking in English. The study examined the effects of two types of instruction on the learning of past -ed allomorphy: Text-to-speech based instruction and one led by a language teacher. The results showed that both groups performed similarly, and both improved in their acquisition of one of the allomorphs: /d/. The study also revealed that there exists a development sequence in the acquisition of regular past tense allomorphy (i.e., /id/ > /d/ > /t/), showing that /id/ was the easiest to acquire while /t/ was the hardest. The results of this study are interesting because they suggest that technology (i.e., TTS) is an efficient, cost-effective and viable solution that can aid learners with their language learning endeavors without exhausting valuable classroom time. So, what if that technology had a targeted interactive component like Intelligent Personal Assistants, for example, where users can talk to it?

This study thus aims to explore the use of Intelligent Personal Assistants (IPAs) as a tool to promote learning, focusing on the acquisition of simple past tense- ed in pronunciation, an under -investigated topic in English morphophonology. The proposed experimental study aims to fill this gap in the literature by investigating whether an IPA, the Alexa App, can assist English learners in improving and developing their phonological awareness and, consequently, their perception and production of the allomorphy that characterizes regular past tense -ed marking in English.

The rationale for using IPAs is based on the results of previous research with IPAs (Dizon, 2017; 2020, Moussalli & Cardoso, 2016; 2020; Underwood, 2017) which indicate that IPAs foster repetition and thus enhance listening and speaking skills. In addition, IPAs can motivate learners to reformulate, self-correct, and persist in using the target language with the help of both implicit (oral) and explicit (transcriptions) feedback received from the IPAs.

**IPAs and Alexa**

The term IPA has been used to refer to Intelligent Personal Assistant (IPA), a voice-controlled service that can complete various functions by orally interacting with its users. IPAs use a combination of three technologies to properly operate and perform the desired tasks. The first is conversational interface which relies on automatic voice recognition tools (ASR) to interpret meaning. The second is personal context awareness, which enables IPAs to understand users' language patterns and specific nuances based on the speaker's pitch, or variation in tone (such as raising intonation for questions, or the use of words that indicate the start of a new sentence as in "Mary" in utterance-initial position). Finally, the third is service delegation, which allows IPAs to integrate with other programs and Apps. For instance, with the additional hardware and sometimes software, IPAs can be used to control and automate smart home

devices. Via these three technologies, IPAs can orally communicate with their users in-human like conversations (Johnson, Brown, & Becker, 2014). An example of a popular IPA is the Alexa App, which will be targeted in this study.

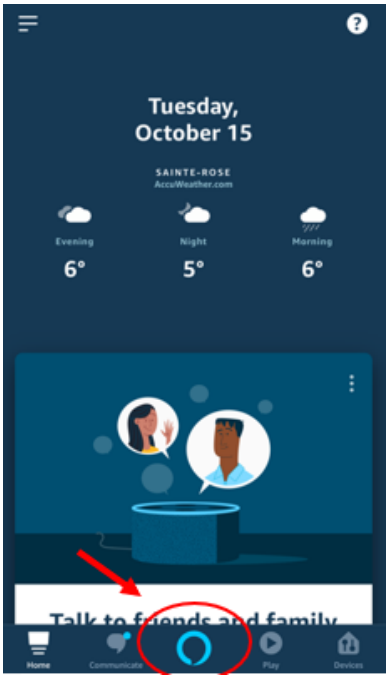
The Alexa App (or simply “Alexa”) is a virtual assistant developed by Amazon that provides immediate response to any oral request. It was first used in the Amazon Echo smart speaker developed by Amazon. Later, the company developed it further into an application that could be downloaded on various mobile devices for at home user experience and on-the-go. The Alexa App provides transcriptions of the interaction with its users (Figure 4.1). It also allows users to enable different categories of Alexa skills that range from education, gaming, and lifestyle to business. It is usually activated on different devices with the wake-word “Alexa”. Some devices require its users to push a button (Figure 4.2) to activate Alexa's listening mod- a feature that allows Alexa to interact with its users by listening to them as they speak in order to respond to the oral request.- while other smart devices may allow users to select their own “wake word”. Another interesting feature is that Alexa is speaker adaptive - i.e., the more a person uses Alexa, the more it adapts to the speech patterns, vocabulary, and personal preferences of that user. However, this adaptation or accuracy improvement can spark privacy concern issues because of the amount of data that is being gathered from users, mainly by listening to their recordings and learning so much about them.

**Figure 0.1**

*Alexa App*


**Figure 0.2**

*Alexa App and command* button



The optimistic (albeit preliminary) findings from previous studies about the adoption of Alexa (Moussalli & Cardoso, 2016; 2020) or other IPAs (Dizon, 2017, Underwood, 2017) as pedagogical tools show that they are easy to use, provide opportunities for input exposure and output practice, and motivate students to learn on their own. Nevertheless, none of the studies above used the Alexa app, a downloadable application that can be used on any mobile device on-the-go, including Amazon Echo. Therefore, in this study, the focus will be on the Alexa App and its potential to contribute to learning.

**Attitudes towards the use of the Alexa App: Student perceptions**

Because all previous studies on IPAs have focused on investigating users' perceptions on the use of *smart speakers* (not the app- although similar in function but different in format and usability), as explained above, it is important to probe learners' attitudes towards the use of Alexa as an independent (not device/speaker-specific) app because it can be downloaded onto any mobile device and used on-the-go. To investigate users' perceptions of the Alexa App in this study, four constructs were adopted, all prevalent in the Computer Assisted Language Learning literature to assess learners' perceptions (Shouma, 2019):

- Learnability: encompasses technology's ability to improve learning skills, enhance subject content, foster engagement, support communication, and develop confidence
- Usability: it is the ease of use and acceptability of a product for a particular class of users carrying out specific tasks in a specific environment
- Motivation: encompasses technology's ability to generate enthusiasm and encouragement toward learning

- Willingness to use: is the positive attitude that users have toward their desire to use and continue using the technology (e.g., to learn other languages or features of a particular language)

In this study, as will be discussed later, participants were asked to evaluate the proposed app (Alexa) via a survey administered at the end of the study. This is an important step in assessing a tool for pedagogical purposes (Cardoso, 2022), particularly in early stages of implementation.

**Simple Past Tense -ed in English**

English simple past -ed has regular (e.g., work/worked) and irregular forms (e.g., go/went) and its negatives and questions are formed with the auxiliary *did* (e.g., she *did* not go to gym last Saturday; *did* he play soccer yesterday?). In affirmative sentences, the regular simple past is represented orthographically by adding the -ed suffix, which has 3 allomorphs: /t/ (e.g., asked); /d/ (e.g., hugged); and /id/ (e.g., hunted). The allomorph /t/ is pronounced with verbs whose base forms end in voiceless obstruents oth er than /t/; /d/ is pronounced with verbs whose base forms end in vowels, sonorants, or voiced obstruents other than /d/; and finally, /id/ is pronounced with verbs whose base forms end in one of the alveolar stops /t/ or /d/.

The simple past tense is particularly difficult to acquire for first and second language learners of English alike (Jia & Fuse, 2007; McDonald & Roussel, 2010) and, for that reason, many researchers have attempted to explain why this is the case. Several reasons and hypotheses have been raised to explain the difficulties that characterize simple past tense acquisition, including: (1) it is infrequent in the L2 classroom, where past -ed comprises only 2% of the verb forms in classroom input to learners (Collins et al., 2009); (2) it is considered redundant (the past meaning is usually expressed elsewhere in the sentence (e.g., *yesterday*, we played chess, where

the past concept is encoded in both the -ed form and lexically by "yesterday" – DeKeyser, 2005; Grause & Coppen, 2015); and (3) it is not perceptually salient (Collins et al., 2009; e.g., allomorphs /t/ and /d/ are often deleted or coarticulated with the following consonant, causing the final stop to be perceived as absent). Due to these reasons, the simple past is acquired late (Jia & Fuse, 2007; Jiang et al., 2011).

Studies that explored the pronunciation of regular past tense found that explicit instruction helps learners improve their production of -ed allomorphs. For example, Mariano (2009) investigated the influence of two types of training (perceptual training versus perceptual training followed by explicit instruction) on the pronunciation of -ed suffixes in sentence reading. The participants were Brazilian learners of English, who were divided into three different groups: control, perceptual training, and perceptual training with instruction. The results showed that only students who were exposed to perceptual training along with explicit instruction improved the pronunciation of the -ed verb endings. Davila (2018) also showed that Nicaraguan English-as-a-foreign language learners acquired the regular past tense allomorphy in a sequence /id/>/d/>/t/ wherein /id/ was easiest to acquire and /t/ the hardest to acquire similarly to the findings of Cardoso (2018). The findings also revealed that the accurate production of the –ed morpheme improved with the learners' level of English proficiency and that explicit instruction was important in improving-ed pronunciation. Similarly, Strachan (2016) showed that English native speakers perceived –ed more accurately than proficient English learners in naturalistic input, and that the more proficient learners were more confident in perceiving it in the target L2 input.

The proposed experimental study aims to investigate whether the selected IPA, the Alexa App, can assist English learners in improving and developing their phonological awareness and,

consequently, their perception and production of the allomorphy that characterizes regular past tense -ed marking in English. To accomplish this goal, it adopts Celce-Murcia et al.'s (2010) Communicative Framework for Teaching Pronunciation.

**The Communicative Framework for Teaching Pronunciation**

According to Celce-Murcia et al. (2010), pronunciation is essential for the achievement of communicative competence and, in order to reach an acceptable level of communicative ability, and assuming that aural perception precedes oral production, learners should be taught via their proposed communicative framework, which consists of five chronological stages (see Gagne, 1985 for a similar approach in the broader field of education):

Stage 1: Develop sound awareness (description and analysis)

Stage 2: Develop listening discrimination (identification and discrimination)

Stage 3: Controlled practice

Stage 4: Guided practice

Stage 5: Communicative practice

Celce-Murcia et al. (2010) explain that learners pass through these stages gradually as they need time to learn the new targeted features and automatize them in their oral production. The authors also advocate that pronunciation practice should extend beyond controlled practice (phrase 3) to reach more creative and communicative exchanges where learners can receive feedback on their pronunciation. In other words, focus on form or the target feature is gradually shifted towards incorporating more meaning.

Specifically, in stage 1 (discrimination and analysis), learners' attention is brought to the associated articulatory features and how these features occur within the language. In the second stage (listening discrimination or aural "perception"), learners' attention is directed towards

identifying and discriminating the target feature (e.g., after listening to it) and providing learners with feedback to help them gradually train their ears and raise their aural knowledge about the target feature. In stage 3 (controlled practice), the focus is on form and accuracy, where learners are encouraged to produce the target feature accurately or intelligibly through repetition and practice. In stage 4: guided practice, the focus is on both fluency and accuracy, where learners practice in contexts devoid of meaning (e.g., producing tongue twisters) or in scripted/guided roleplay activities. In the final stage (communicative practice), learners are required to use the just-learned feature in genuine interactions such as in tasks that require negotiation of meaning, where focus is placed on both accuracy (in terms of intelligibility) and content. Examples of these activities include (unscripted) roleplay, storytelling, and interviews. In sum, after exposing students to the target feature (and the rules that govern its production) and providing them with sufficient opportunities to produce it along with corrective form-focused feedback, students will be able to shift their explicit knowledge to spontaneous, automatized production (Reed 2012; Yang & Lyster 2010).

**This Study**

The framework for pronunciation instruction set forth by Celce-Murcia et al (2010) assumes that, for successful L2 pronunciation pedagogy, learners should be instructed in a way in which their phonological acquisition starts with awareness raising (stage 1), and proceeds via the development of perception (discrimination abilities- stage 2), controlled (stage 3) and guided (stage 4) oral production, towards a more spontaneous and automatized use of the target feature (stage 5). Accordingly, this experimental study employed in its design the first *four* stages recommended by the authors. The study examined the participants' awareness to the sounds involved in past -ed marking, their aural perception (or ability to discriminate among the three

allomorphs) and oral production. In addition, it also examined the participants' attitudes toward the pedagogical use of IPAs in their learning of English -ed morphophonemics. The study is guided by the following research questions:

1. Will the pedagogical use of a commercial IPA (Alexa) help English learners acquire English past-ed allomorphy in terms of:
    a) Phonological/sound awareness (Stage 1)
    b) Perception (or phonemic discrimination; Stage 2)
    c) Production (Stages 3 and 4)
2. What are learners' attitudes towards the pedagogical use of the Alexa App for the learning of the target English morphophonemics?

It is hypothesized that a focus on listening and oral production training via the Alexa App will lead to an improvement in phonological awareness, aural perception, and oral production of past tense –ed forms. In addition, it is predicted that participants will view the technology and related pedagogy positively, as observed in previous studies.

## Method

### Participants

Participants (N=18, 9 males and 9 females) from different language backgrounds were recruited for the study. They were randomly divided into two groups: the Alexa group and the non-Alexa group, each consisting of 9 learners. They were international ESL students recruited from a university or college, and their oral proficiencies in English were from low intermediate to advanced, based on their self-assessment (the call for participants included that requirement), and the researcher's assessment of their responses to the language background questionnaire (Appendix C). Their ages are from 18 to 30 years old.

## Material and Design

This study consisted of five main phases: (1) Testing phase 1 (pre-test), (2) Explicit -ed instruction phase, (3) App familiarization phase (how to use the Alexa App - Alexa group only), (4) Practice phase, and (5) Testing phase 2 (post-test) which will be explained thoroughly in the forthcoming section.

### *Instruments*

**Background questionnaire and consent form**. At the beginning of the experiment, the participants were asked to fill out a consent form (Appendix I) and a language background questionnaire (Appendix C) to gather demographic data and other relevant information such as their previous language learning experience. The questionnaire consisted of 11 questions in English about the participants' first languages, their language learning history, and their knowledge and use of technology.

**Pre-tests and Post-tests.** Three sets of tasks were used in the pre-test and the post-test, each consisting of one or two tests to triangulate data collection and to examine different types of phonological knowledge: phonological awareness (n=1, but divided into two tasks), aural perception (n=2), and oral production (=2), in that order (Figure 4.3).

Set 1, an awareness test (Appendix J), is based on a single survey that asked the participants to answer questions about their knowledge about past -ed pronunciation using survey completion (part 1) and think-aloud protocols (part 2). The two parts took place concurrently, i.e., the think-aloud (section 2) took place during and after the survey completion (concurrent and retrospective think-aloud; Bowels, 2010). The think-aloud was adopted to gain further information concerning the participants' understanding of the simple past tense -ed allomorphy. For the survey completion (section 1), the participants were asked to respond to statements such

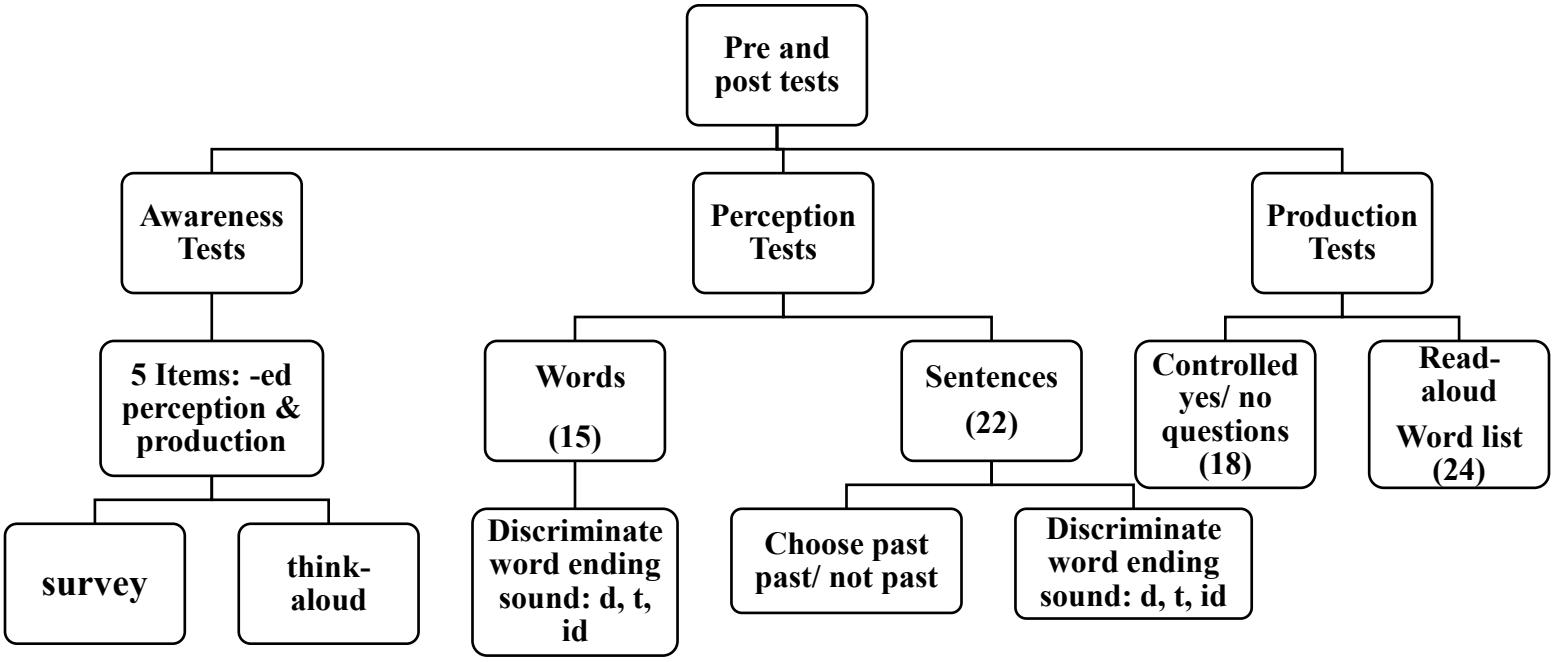
as “past -ed is pronounced the same in verbs such as *walked*, *lived* and *invited*”, using a 9-point Likert scale (ranging from 1=strongly disagree to 9=strongly agree). This test aimed to quantify participants responses to 5 statements about their knowledge of the simple past tense -ed allomorphy (e.g., “the past -ed is pronounced the same in verbs such as walked, lived and invited”). All questions were pilot-tested by a group of 5 participants.

Set 2, the two perception tasks (listening discrimination) consisted of identifying the past –ed sound in: (a) a set of 15 words (Appendix K) with allomorphic discrimination (/t/ /d/ or /id/), and (b) a set of 22 short sentences (Appendix L), divided into two parts (both adapted from Cardoso, 2018). In the first part, participants were asked to decide whether each sentence they heard was in the past or not (past, not past); if the sentence was in the past, they were then asked, in the second part, to discriminate among the three verb ending options (/t/ /d/ or /id/).

Set 3, the two production tasks (both adapted from Cardoso, 2018) consisted of the reading aloud of a word-list containing 24 verbs inflected for past -ed (Appendix M) and a controlled mock role-play questions (Appendix N) about a past event that targeted yes/no answers to questions such as “did Kevin work during his vacation?”. The testing materials for pretests and posttests were similar in content, but not identical, and included highly frequent regular past tense forms. Furthermore, the target -ed forms were equally distributed in all tasks.

**Figure 4.3**

Perception and Production Tests: Pre and Post Tests




***Survey***

To assess the participants' attitudes towards the pedagogical use of the target IPA, they were asked to complete a survey at the end of the experiment, after the post-tests. The survey (Appendix O) is a four-section survey (n=28) with each section consisting of 7 statements evaluated using a 9-point Likert scale (ranging from 1=strongly disagree to 9=strongly agree). These items aim to quantify their responses to several statements about their attitudes towards their pedagogical experience using the target IPA. Specifically, the four sections target themes that are prevalent in CALL literature (Shouma, 2019): learnability (Diemer et al.,2013 ; for example, using the Alexa app helped me improve my pronunciation of past tense -ed in English), usability (Lawrence, 2016; for example, I find it is easy to use the Alexa App), motivation ( Dörnyei, 2009 ; for example, using Alexa App motivates me to learn about how words are pronounced in English), and willingness to use the IPA (Cardoso, 2011; for example, I would

like to continue to use Alexa App to practice speaking in English). As customary in survey studies, the methodology for designing and validating its items followed standard research methods protocols (as such, the target items were checked for internal consistency, and were later revised based on pilot-testing and users' feedback; for details, see Liakin et al., 2015). Also, in order to measure scale reliability, Cronbach's Alpha was calculated and, as will be described later, the values indicated satisfactory internal consistency between the items for each of the adopted themes.

***Interviews***

A semi-structured oral interview was conducted with each participant for an average of 30 minutes after the post-test to obtain more in-depth insights about the participants' pedagogical experience using the Alexa App. The questions reflected the goals of the study so that we could have a better understanding of the learner's knowledge of the -ed allomorphy and their pedagogical experience using the Alexa App. Example of questions included: 'How did you like your experience with the Alexa App?'; 'Did you understand Alexa? Do you hear anything different in the pronunciations of some verbs in the past? Explain. Do you know if there's a different between the past tenses in WALK, LIVE and INVITE? Do you think Alexa helped you pronounce the past tense better?

***Raters***

Two English second language teachers scored (1= correct, 0= wrong) all produced speech samples extracted from the participants, which consisted of 24 read-aloud words and 18 controlled yes-no sentences (example, 'yes he walked to the airport", "no, he visited his parents"). The average length of the sentences were 5.3 words in length, targeting the simple past tense in English in regular verbs. Both the read-aloud word list and controlled yes-no questions

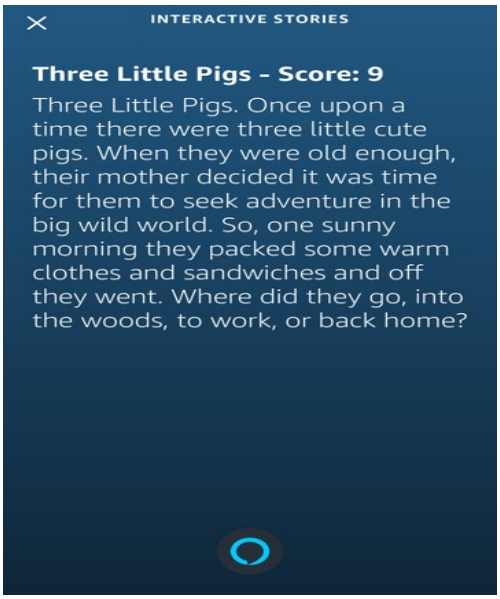
included highly frequent verbs inflected for past -ed, i.e., within the 1,000 most frequently used verbs in English. The rationale behind the inclusion of these raters is to compare their scoring to be as consistent and as accurate as possible.

***Practice Material***

The practice material consisted of a list of four sets of 10 verbs ending with -ed for participants to practice the perception and production of English past -ed (Appendix P). The list contained highly frequent regular past tense forms that are different from those used in the tests. All participants were provided with this list. The Alexa-group practiced with the Alexa App, using two Alexa skills: (a) "Simon says", in which participants asked the app to repeat the produced target words, and (b) "How do you pronounce P.L.A.Y.E.D." to help participants listen to how the target word is produced. The non-Alexa group, on the other hand, were asked to practice the same list with friends, their parents, or by themselves (similar to what students are asked to do when they complete homework assignments). In addition to the above practice, the Alexa group were asked to interact with Alexa by, for example, listening to "interactive stories" (see Figure 4.4) - an Alexa skill that allows participants to listen to stories and interact with the app by answering questions. To emulate the treatment received by the Alexa group, the non-Alexa group were asked to listen to pre-selected five-minute "TED Talks" and orally summarize the story.

**Figure 4.4**

*Interactive Stories*




**Procedure**

This study consisted of five main phases: (1) Testing phase 1 (pre-test), (2) Explicit -ed instruction phase, (3) App familiarization phase (how to use the Alexa App - Alexa group only), (4) Practice phase, and (5) Testing phase 2 (post-test) (see Table 4.1 below).

Before the pre-test, participants first provided consent and completed a written questionnaire on their language learning history. Soon after, they completed the six pre-tests (three tests for each aspect of pronunciation development). After the pre-test, participants were randomly assigned to their groups (Alexa or non-Alexa groups). The participants then underwent explicit -ed instruction, familiarization with the Alexa App (Alexa Group only), and the practice phase (30-45 minutes per session- see next section). After the pedagogical intervention, which lasted approximately four weeks (4.5h in total), the participants completed the post-test. In the end, all participants were invited to participate in a 30 minute-semi-structured interview (Appendix Q). The participants in the Alexa group were asked about their experience and attitudes towards the use of the adopted IPA as a pedagogical tool and their understanding of the

target feature, whereas the participants in the non-Alexa group were interviewed about how they completed their practice material and their understanding of the target phonological feature.

**Table 4. 1**

*Overview of design study*

| Alexa | Non-Alexa |
|---|---|
| Week 1 | |
| Language Questionnaire, pre-test phase & explicit -ed instruction phase | |
| App familiarization phase | N/A |
| Practice phase as homework | |
| Weeks 2-4 | |
| Practice phase as homework | |
| End of week 4 | |
| Post- test phase & interview | |
| Alexa survey | N/A |

***Testing phase***

There were two testing phases: The pre-test and the post-test. The testing materials for both the pre-test and the post-test are similar and include highly frequent verbs inflected for past -ed, i.e., within the 1000 most frequently-used verbs in English to avoid frequency effects (Ellis, 2002). Both pre-tests and post-tests consisted of three sets dedicated to assessing the participants' development in awareness (set 1), perception (set 2), and production (set 3).

***Explicit -ed Instruction phase***

After the completion of the pre-test, participants were randomly assigned to two groups: the Alexa group and the non-Alexa group- where each even-number participant went into one group and odd numbered one went into the other. Both groups were provided with explicit teaching by the same native teacher, in the same manner, about the past -ed inflection and its

pronunciation: /id/ (as -ed in hunted); /d/ (hugged); /t/ (asked). As typical in language classrooms, in addition to in- person class practice and feedback provided by the teacher, the teacher assigned controlled practice exercises as homework (outside the class environment) so that participants in both groups can practice the production and perception of the ed- allomorphs: While the Alexa group practiced with the Alexa App, the non-Alexa group did the same with friends or classmates, their parents, or by themselves. The same exercises were given to both groups (Appendix P).

***App Familiarization Phase***

After the pre-tests and the explicit -ed instruction, participants in the Alexa group were asked to download the Alexa App on their phones and were taught how to use it. The participants were provided with a pre-established set of questions (n=30) created by the researcher , and they were encouraged to interact with the Alexa App during the familiarization session for approximately 15 minutes. The participants were encouraged to use the different Alexa skills such as “interactive stories”, which allowed participants to listen to stories and interact with the Alexa App by answering questions.

***Practice Phase***

The two groups of participants were provided with a list of verbs ending with -ed to practice the perception and production of the simple past tense for approximately 15 minutes, three times a week. Participants in the Alexa group were asked to interact with Alexa using the list of verbs provided to them, the pre-established set of requests, as well as their own personal questions (see practice material section in Appendix R). They were asked to use the Alexa App to listen to fables containing the target feature (e.g., stories about past events) and interact with the app by answering questions through the “interactive stories” skill. The non-Alexa group were provided with the same list of verbs and some guided activities; however, they were asked

to practice the perception and production of past -ed with their classmates, their parents or alone. In addition. they were asked to listen to pre-selected “TED Talks”, which were no longer than five minutes, and then summarize the talk orally.

The practice material, which were assigned to both groups as homework, lasted for approximately 4.5 hours in total over a period of three weeks. The reason the non-Alexa group were not provided with specific tools was to create an authentic learning setting where learners are given the option of choosing their own method of learning in the absence of the Alexa App. For example, students could ask a fluent speaker of English or a teacher to help them practice. Both groups were asked to respect the practice time-limit provided. They were also asked about their time-on-task during the oral interview.

## Analysis

A multimethod design was used in this study. Quantitative data were analyzed from pre-test and post-test with appropriate methods for the analysis of quantitative data (between and within mixed ANOVAs). For instance, all pre- and post-test scores were scored by two English second language teachers as 1 point for correct and 0 for wrong for the perception (or phonemic discrimination; Stage 2) and production tests (Stages 3 and 4). Means and standard deviations for the 9-point Likert rating scale and phonological awareness scales were calculated. Finally, qualitative data from the think-aloud and interviews were transcribed and analyzed according to the coding methods proposed by Saldaña (2009), which recommends that the analysis should move through different coding cycles: from creating codes to general categories in the first cycle, to narrowing it down to more specific themes and concepts in the second cycle. This coding method helped organize the participants’ intended meaning for analysis. Also, to ensure homogeneity among the two judges who scored the participants’ production, Cohen's kappa

coefficient was calculated to determine inter-rater reliability for the scoring of the production tasks: the read-aloud word list and controlled yes-no questions. The result suggested a moderate level of reliability, Cohen's kappa $\kappa = 0.561$. In cases where the two raters did not agree, they discussed and agreed on what the answer should be.

## Results

To answer the first research question (will the pedagogical use of a commercial IPA, Alexa, help English learners acquire English past-ed allomorphy?) quantitative analysis were performed to examine participants' development across three levels (or stages, as per Celce-Murcia et al.' 2010): phonological awareness (set 1), aural perception (set 2), and oral production (set 3).

### Quantitative Results

Starting with the results of the phonological awareness test, set 1, the results of a Wilcoxon signed rank test revealed no significant difference in both groups' performance over time (see Table 4.2) but the descriptive results (means) revealed that participants in the Alexa group improved between the pre- and post-test for the first test (survey). For example, the means for "ed-*accurate*" statements (e.g., "-ed in *kissed* and *jumped* sound the same") increased for the Alexa group from M=3.75 to M=6.61, while for the non-Alexa group, the means dropped slightly. Contrastively, for "ed-*inaccurate*" statements (e.g., -ed is pronounced the same in *walked*, *lived* and *invited*", the means decreased from M=5.25 to M=3.75, as hypothesized for the Alexa group, but increased for the non-Alexa group from M=3.2 to M=3.4 (see Table 4.3).

**Table 4.2**
*Awareness test: Wilcoxon signed rank test*

| | ALEXA | Non-ALEXA |
|---|---|---|

| | | z | Asymp. Sig. (2-tailed) | z | Asymp. Sig. (2-tailed) |
|---|---|---|---|---|---|
| ed-accurate | I know how to pronounce the past -ed in English. | -1.594 | .111 | .000 | 1.000 |
| | In English, the past -ed in **KISSED** and **JUMPED** sound the same | -1.219 | .223 | -1.342 | .180 |
| | In English, the past -ed in **PRINTED** and **DREAMED** sound the same | -.339 | .735 | -.184 | .854 |
| ed-inaccurate | In English, the past -ed in **JUMPED** and **DREAMED** sound the same | -.524 | .600 | -1.069 | .285 |
| | In English, the past -ed is pronounced the same in verbs such as **WALKED, LIVED** and **INVITED** | -1.069 | .285 | -.535 | .593 |

**Table 4.3**

*Awareness test: Means and standard deviation*

| | | ALEXA | | | | Non-ALEXA | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Pretest | | Post-test | | Pretest | | Post-test | |
| | | M /9 | SD | M /9 | SD | M /9 | SD | M /9 | SD |
| ed-accurate | I know how to pronounce the past -ed in English. | 5.88 | 0.58 | 6.88 | 0.35 | 5.5 | 0.82 | 5.5 | 0.7 |
| | In English, the past -ed in **KISSED** and **JUMPED** sound the same | 3.75 | 1.15 | 6.61 | 3.12 | 3 | 0.83 | 2.7 | 0.79 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| ed-inaccurate | In English, the past -ed in **PRINTED** and **DREAMED** sound the same | 4.88 | 1.25 | 3.63 | 1.29 | 3.1 | 0.98 | 2.9 | 0.94 |
| | In English, the past -ed in **JUMPED** and **DREAMED** sound the same | 5.75 | 0.94 | 3.56 | 1.21 | 5.4 | 1.19 | 6.8 | 0.96 |
| | In English, the past -ed is pronounced the same in verbs such as **WALKED, LIVED** and **INVITED** | 5.25 | 1.03 | 3.75 | 1.15 | 3.2 | 0.85 | 3.4 | 0.92 |

Consider now the results of set 2 (the perception test), which consisted of two parts: the word and sentence discrimination test. The results revealed that there were no significant improvements in participants' performance from pre-test to post-test (see tables 4.4 and 4.5). After scoring both tests with 1 for correct answers and 0 for incorrect answers, the analysis of a mixed ANOVA for the first part of the test, the word discrimination task, revealed that there were no significant main effects between the pre and post-test for the word perception task, $F(1,16) = .119$, $p = .73$, $\eta_p^2 = .007$ and no significant interaction effect between the groups over the pre and post-test $F(1,16) = .007$, $p = .932$, $\eta_p^2 < .001$. Based on these findings, we can conclude that there was no difference between the groups in the word discrimination task over time, and that the pre- and post-test scores between both groups were similar in the word perception task $F(1,16) = 0.018$, $p = 0.896$, $\eta_p^2 = .001$.

Similarly, the results for the second part of the perception task, the sentence discrimination task, revealed non-significant main effects between the pre and post-test $F(1,16) = 2.47$, $p = .136$, $\eta p^2 = .134$ and no significant interaction effect between the groups during the pre and post-test $F(1,16) = .254$, $p = .621$, $\eta p^2 = .016$. These findings indicate that there was no difference between the groups in the sentence discrimination task over time, and that the groups

did not significantly change over the duration of the experiment $F(1,16)=.000$, $p = 0.990$, $\eta_p^2 <.001$. Interestingly, when the three allomorphs were analyzed separately, the results showed some improvement trends although not statistically, especially for /t/ in the Alexa group: the means for this allomorph more than doubled, with an increase in perception from M=4.16 (pre-test) to M=9.16 (post-test).

**Table 4.4**

*Perception tests: Means and standard deviation*

| | **Alexa** | | **Non-Alexa** | |
|---|---|---|---|---|
| | **M /10** | **SD** | **M /10** | **SD** |
| Word discrimination: pretest | 7.7 | 1.99 | 7.8 | 4.21 |
| Word discrimination post-test | 8.5 | 2.811 | 7.6 | 2.59 |
| Sentence discrimination: pretest | 5.3 | 3.77 | 5.5 | 5.11 |
| Sentence discrimination: post-test | 6.1 | 5.52 | 5.9 | 3.44 |

**Table 4.5**

*Perception tests: Means and standard deviation per allomorph*

| | Word discrimination | | | | Sentence discrimination | | | |
|---|---|---|---|---|---|---|---|---|
| | **Alexa** | | **Non-Alexa** | | **Alexa** | | **Non-Alexa** | |
| | **Pre-test** | **Post test** | **Pre-test** | **Post test** | **Pre-test** | **Post test** | **Pre-test** | **Post test** |
| /Id/ | 9.16 (1.44) | 8.33 (.72) | 8.66 (.57) | 8.33 (.57) | 6.56 (2.77) | 5.94 (1.57) | 6.75 (.95) | 7.5 (.57) |
| /d/ | 7.5 (2.5) | 8 (.72) | 9.16 (0) | 10 (2.08) | 4.06 (3.12) | 5.94 (2.36) | 5.5 (1.29) | 6.75 (.95) |
| /t/ | 4.16 (1.44) | 9.16 (.72) | 5.3 (4.6) | 7.3 (1.15) | 5 (1.76) | 5.63 (1.25) | 5.75 (3.20) | 5.25 (2.87) |

*Mean score out of 10; numbers between parenthesis indicate Standard Deviation Error.

Finally, the results of set 3 (the production test), which consisted of two parts (read-aloud word list and controlled yes-no questions), revealed that there were no significant improvements in participants' performance from pre-test to post-test (see tables 4.6 and 4.7). After scoring 1 for every correct pronunciation of the -ed allomorphs, and 0 for an incorrect production, the

results of the mixed ANOVA for the read-aloud word list revealed non-significant main effect between the pre and post-test $F(1,16)=3.72$, $p=.072$, $\eta p^2 = .189$ and no significant interaction effect between the groups and the pre and post-test $F(1,16)=1.04$, $p=.323$ , $\eta p^2 =.016$. Based on these findings, we can conclude that there was no difference between the groups in the read-aloud word list task over time, and that the two groups did not significantly change over time $F(1,16)=.002$, $p = 0.966$, $\eta_p^2 <.001$.

Interestingly, the results of the second production task, the controlled yes/no questions, revealed significant main effect between the pre and post-test $F(1,16)=4.734$, $p=.045$, $\eta p^2 = .228$ but no significant interaction effect between the groups and the pre and post-test $F(1,16)=1.935$, $p=.183$ , $\eta p^2=.108$. These findings suggest that all participants improved significantly between the pre-test and the post-test, but there was no difference between the groups (i.e., the groups did not change differently over time; they changed in the same way); $F(1,16)=1.59$, $p = .226$, $\eta_p^2 =.090$.

It is worth noting that the results of Levene's test of equality of error variances revealed that all values were >.05 which means that the variances are homogeneous for all levels of repeated measures.

**Table 4.6**

*Production tests: Means and standard deviation*

| | **Alexa** | | **Non-Alexa** | |
|---|---|---|---|---|
| | **M /10** | **SD** | **M /10** | **SD** |
| Read-aloud word: pretest | 7.8 | 3.44 | 8.1 | 2.71 |
| Read-aloud word post-test | 8.5 | 3.50 | 8.3 | 3.26 |
| Controlled yes-no questions: pretest | 4.5 | 5.86 | 4 | 6.19 |
| Controlled yes-no questions : post-test | 7.2 | 4.25 | 4.6 | 5.83 |

**Table 4.7**

*Production tests: Means and standard deviation per allomorph*

| | Production words | | | | Production sentences | | | |
|---|---|---|---|---|---|---|---|---|
| | **Alexa** | | **Non-Alexa** | | **Alexa** | | **Non-Alexa** | |
| | **Pre-test** | **Post test** | **Pre-test** | **Post test** | **Pre-test** | **Post test** | **Pre-test** | **Post test** |
| /Id/ | 8.75 (1.18) | 8.95 (.51) | 8.5 (1.04) | 9 (.89) | 4.79 (1.46) | 8.12 (1.61) | 4.33 (1.21) | 4.25 (1.25) |
| /d/ | 7.08 (1.70) | 8.12 (1.04) | 8.3 (1.50) | 8.5 (1.37) | 5.2 (2) | 7.65 (1.69) | 4.66 (1.03) | 5 (1.92) |
| /t/ | 5.62 (1.72) | 7.29 (2) | 5.8 (1.72) | 5.8 (3.25) | 3.54 (1.46) | 6.25 (.79) | 2.83 (1.16) | 4.16 (1.32) |

*Mean score out of 10; numbers between parenthesis indicate Standard Deviation.

To answer the second research question (what are learners' attitudes towards the pedagogical use of the Alexa App for the learning of the target English morphophonemics?), which examined learners' attitudes towards the pedagogical use of the Alexa App for the learning of the target English morphophonemics, means and standard deviations were calculated from the ratings of the 28-item, four-theme survey. As illustrated in Table 4.8, the results of this survey (1 = strongly disagree and 9 = strongly agree) revealed that, from a learners' perceptions standpoint, Alexa has great potential as a learning tool (learnability: M=7.07 /9), it has high usability scores (usability: M=6.77 /9), it motivates the participants to learn and explore the language (motivation (M=7.3 /9), and it is likely to spark their willingness to continue to use the IPA in their future language learning endeavours (willingness to use: M=7.73 /9).

**Table 4.8**

*Four-theme survey: Means and standard deviation*

| | Mean/9* | SD | Cronbach's Alpha |
|---|---|---|---|
| Learnability (n=7) | 7.07 | .36 | .89 |
| Usability (n=7) | 6.77 | .39 | .82 |
| Motivation (n=7) | 7.36 | .41 | .92 |
| Willingness to use (n=7) | 7.73 | .47 | .94 |

*Note*. survey endpoints were 1 = strongly DISAGREE and 9 = strongly AGREE

**Qualitative Results**

To answer the second research question (what are learners' attitudes towards the pedagogical use of Alexa for the learning of the target English morphophonemics?), the analysis was built up by combining the findings of the think aloud and interview data. These data were transcribed and analyzed qualitatively following the coding methods proposed by Saldaña (2009), as discussed earlier. The findings were analyzed based on the four research themes or constructs: learnability, usability, motivation, and willingness to use.

Starting with learnability, the learners displayed a positive attitude toward using the Alexa App for the learning of the target English morphophonemics. For example, an analysis of the post-test think-aloud task showed that participants gained sound awareness of the allomorphs of English past -ed, as explained by some learners: "before [the practice], I think when I say past tense, it's easy to me because I don't have anything about this [the three different -ed allomorphs]; but now I worry and confusing but yes easy for me to say", and "Before I didn't hear [the three different -ed allomorphs], now yes". In addition, the participants felt more confident about the three different -ed allomorphs, and they understood that words like *walked*, *lived* and *invited* are pronounced differently: "Yes, I think so, they are different sounds, so they are all of them different sounds"; "Verb different change ed sound". Moreover, the analysis of the interview revealed that the learners found it harder to perceive the different /t/, /d/, [ɪd]

allomorphs: “Hearing the past ed, I am confusing”. “Not easy [to hear the difference] I need thinking about what sound the last sound”. In addition, some participants commented that they have had difficulties discriminating between /t/ and /d/, but not /id/: “A bit difficult to hear, t and d; id sound is clearly clearer”; “Listen [is harder] because is very ended when people say too quietly I can hear it, but the t, d, I hear all d. I can’t hear t, d. Harder in sentence than words”; “I saw the difference looking for your mouth, maybe sometimes could be harder because I can’t understand the end of the word. It’s more clear for id because it is a long sound”. And lastly, the learners stated that it was easier to produce the target /t/, /d/, [ɪd] allomorphs than to perceive them, with /t/ and /d/ being more challenging than /id/: “For me, it’s easy to say, I think it’s difficult to hear sometimes, just part normally the finish”; “For example I don’t know how can I pronounce kissed with t or d; yeah, hard”; “ No [can’t hear the difference], only when I say, I hear difference”; “Speak is easy, hear is a more difficult [than to speak], something invited is okay but t and d is hard”, “Hear is more difficult [than to speak], I coufused d and t sound, when I hear”, “When I hear some sounds, I am a little nervous. It’s not easy but so it’s okay to say [produce]”.

Regarding usability, the majority of the learners highlighted in the interviews the convenience of using the Alexa App “Alexa is good to learn English”. They also mentioned that Alexa was a fun and easy tool to use outside the language classroom as a conversational partner (e.g., where one can practice both listening and speaking skills) “because sometimes you don’t have other person to speak, so Alexa is a tool for this when you are alone”.

Concerning motivation, eight of the learners explained that the app’s use motivated them to interact in the target language and provided visual but implicit feedback to help them understand and self-correct “good practice English for speaking and listening”, “when you

pronounce wrong the application repeat other sentence, and I look at word in the App", "some other time I only understand 50%, but the application write; it helped". In sum, learners emphasized their motivation and interest in using the Alexa App to learn the target morpho-phonological phenomenon.

Finally, all nine learners expressed their willingness to use the Alexa App in the future to help improve their English listening and speaking skills "of course [I plan to use Alexa] to learn English more, listening can help me so I am not good at listening".

In sum, the quantitative analysis revealed some non-statistically-significant improvements in the groups' performance based on the phonological awareness tests, and some improvements for the /t/ allomorph for the Alexa group in perception (via the word discrimination test). However, the results of a mixed ANOVA revealed that there were no differences in the groups' performance over time (i.e., between the pre-test and the post-test) for the perception test (the word and sentence discrimination test), and for the production test (read-aloud of a word list and controlled yes-no questions).

Validating the above results, the qualitative analysis of the interviews and the think-aloud indicated that the participants in the Alexa group gained a certain level of phonological awareness to past -ed allomorphy. In addition, the qualitative results also revealed that some participants found it easier to produce than to perceive the different -ed allomorphs, particularly /t/, /d/. Furthermore, the qualitative analysis also showed that Alexa has great potential for learning, it is highly usable, it increased the learners' willingness to use the app in their future language learning endeavours as it motivated the participants to interact with it using the target language.

## Discussion and Concluding Remarks

This study investigated the ability of the Alexa App, an IPA, to assist ESL learners in developing their phonological awareness, perception, and production of the allomorphy in regular past tense marking in English (e.g., talk/t/, play/d/ and add[ɪd]). It addressed the following questions: (1) Will the pedagogical use of a commercial IPA (Alexa) help English learners acquire English past-ed allomorphy in terms of phonological awareness, perception, and production? And (2) What are learners' attitudes towards the pedagogical use of the Alexa App for the learning of the target English morphophonemics?

To answer the first research question, a quantitative analysis was conducted. The results of the quantitative data revealed improvements in phonological awareness only for the group that interacted with Alexa although this improvement was not significant. Overall, there were no differences in the two groups' performance over time for both perception (i.e., word discrimination and sentence discrimination tests) and production (read-aloud of a word list and a controlled yes-no questions).

Focusing on the results regarding phonological awareness, the participants improved in their ability to recognize the three -ed allomorphs, thus validating Celce-Murcia et al.'s (2010) framework by showing that learners develop their pronunciation in a systematic way, going from their sound awareness and then moving towards the other stages of acquisition. This means that the first stage of Celce-Murcia et al.'s (2010) framework has been supported with regards to the development of sound awareness. Similar findings were also observed in a study by Khademi and Cardoso (in print), which showed that the pedagogical use of Google Translate speech recognition and text synthesizer (TTS) successfully raised the participants' awareness of English past -ed allomorphs. We can thus conclude that IPAs seem to have the potential to help

participants gain sound awareness, especially regarding the development of English -ed morphophonemics.

Regarding perception, the results of the quantitative data for the perception tests revealed that the only learnable allomorph was /t/ for the word discrimination task for the Alexa group. This group's score for /t/ more than doubled in the posttest, considerably higher than the non-Alexa group. Nevertheless, the results of the ANOVA showed no difference in perception between the Alexa group and the non-Alexa group from the pre-tests to the post-tests in both tasks, possibly because of the high levels of standard deviations observed for both groups. It is possible that the results were not significant because the participants might have reached a ceiling for /id/ (8.7- 9.2 out of a possible 10 in the pre-test and /d/ 7.5-9.2 out of a possible 10 in the pre-test. As such, this pattern can be attributed to the fact that the learners' awareness of this allomorph was already high at the outset for both groups (for similar claims, see Khademi & Cardoso in press; Rifkin, 2005).

Finally, concerning production, the results of the quantitative data revealed no difference between the Alexa group and the non-Alexa group from the pre-tests to the post-tests in word production; however, both groups improved significantly from the pre-test to the post-test for the controlled yes-no question, but with no difference between the groups. This improvement in performance with the controlled yes-no question task could be due to the nature of the task, which motivates interaction. The only task that resembled in nature the interaction afforded by the Alexa App was the controlled yes-no questions, in the production test. In other words, interacting with the researcher during the controlled yes-no questions emulated both groups' practice phase (whether it was with Alexa or with friends and/or parents), which could have triggered their learning and perhaps generated these results.

Overall, it can be concluded that there was lack of significant improvement between groups which could be due to the duration of the experiment, which only lasted four weeks. According to a previous study on the acquisition of -ed allomorphy via TTS, Cardoso (2018) concluded that participants need extensive spaced practice to fully acquire -ed allomorphy, particularly when learning implicitly in an autonomous setting because this type of learning can be affected by loss of motivation. Nevertheless, we can also conclude that participants did better on the production tasks than on the perception tasks, an observation that was corroborated by the computation of the means and standard deviation since the means were slightly higher for those of the production tasks than those of the perception tasks. This generalization can also be supported by the qualitative results from the think-aloud and interviews, which will be explained next.

To answer the second research question, which examined learners' attitudes towards the pedagogical use of Alexa for the learning of the target English morphophonemics, quantitative and qualitative analysis were conducted. Data obtained from the four theme-based analysis revealed the following:

In terms of learnability, which evaluated Alexa App's ability to foster learners' engagement and communication to improve learning, the findings revealed that the Alexa App has potential as a learning tool, as it can serve as an interactive partner (just like friends or classmates), or as a source of second language input. More importantly, Alexa's use promoted autonomous learning, and supported language development in similar ways to instructed language learning, corroborating previous research by Diemer et al. (2013) and Dizon (2020). For instance, the learners in the Alexa group (unlike the non-Alexa group) noted that they found it easier to produce the target /t/, /d/, /ɪd/ than to perceive them, with /t/ and /d/ being more

challenging to produce than the /id/ sounds. One possible explanation may be due to orthography, as /id/ is the most transparent of the three allomorphs with a clear spelling-to-sound correspondence (Delatorre, 2010). Both production tasks used in the pre-test and post-test required participants in both groups to read words aloud in isolation and in controlled settings, which could have caused the orthographic effect. Additionally, the learners expressed that it was easier for them to hear the /id/ sound than to hear the /t/ or /d/ sounds. This could be explained by the perceptual salience hypothesis, which predicts that sounds and features that are prosodically salient are acquired earlier and more easily than those that are less salient. According to Klein et al. (2004), a language learner is able to perceive and produce a syllabic grammatical suffix such /id/ (hunted) more accurately than a non-syllabic allomorph such as /d/ (hugged) and /t/ (asked) because a syllable such as /id/, which has a vowel, is more perceptually salient than a consonant (or cluster of consonants) such as the final stops /t/ and /d/. This was articulated by one of the participants, who claimed that the /id/ form was longer than the others and, consequently, the easiest to perceive. These findings are supported by previous research that show that perceptual salience plays an important role in the learning of L2 features (e.g., Collins et al., 2009; Goldschneider & DeKeyser, 2001; Klein et al., 2004; Solt et al., 2003).

Regarding usability, which looked at learners' level of enjoyment and how user-friendly the Alexa App is, the analysis indicated that the application is user friendly and easily accessible, confirming previous findings in the general IPA literature (e.g., Dizon 2016, 2020; Underwood 2017).

Concerning motivation, which measured learner enthusiasm and its overall effects on learning, the results uncovered that the pedagogical use of Alexa is perceived by the learners to motivate them to interact with it, learn and explore the target language.

Finally, considering willingness to use, which assessed learners' perception towards using the Alexa App in their future learning endeavors, the data showed that the Alexa App contributed to increasing their willingness to use it as a conversational partner to practice both listening and speaking skills.

Overall, IPAs and/or their associated apps (the Alexa App in this study) have the potential to engage learners in a continuous process of interactions to process and produce the target language, as suggested by interactionist approaches to second language development (e.g., Long, 1996).

## Limitations and Future Directions

The current study contributes to the CALL literature by demonstrating that IPAs are valuable pedagogical tools that can extend the reach of the classroom by allowing language learners to autonomously improve aspects of their second language phonological development (e.g., awareness of past tense marking (/id/, /t/, /d/). The participants confirmed that Alexa is a promising tool for learning, and it motivated them to use the target language, as attested in qualitative data (interviews), thus corroborating findings highlighting the potential of IPAs to support second language development (Dizon, 2020).

Nevertheless, there are several limitations that need to be addressed in future research. The first is the short duration of the study, which lasted four weeks. In addition, the participants were given between 30 to 45 minutes of time (three times a week) to interact with the IPA, a total of 4.5 hours over 4 weeks. Perhaps providing them with longer period to practice would have yielded more robust results considering the three stages of phonological knowledge considered in the study. The second limitation is the low number of participants. Despite the evidence that the Alexa App can help learners improve their phonological awareness, it is not

clear whether these findings are generalizable to the larger population of English learners. Finally, the third limitation is the nature and complexity of some of the tasks used in the tests, which did not correspond with those used for the interaction with the IPA. During their interaction with the IPA, the participants were asked to use specific skills that allowed them to mainly listen to and repeat and/or produce the targeted features with the help of the IPA. However, the second exercise of the perception test for both pre-test and the post-test (sentence discrimination test) was different in nature and in difficulty level. It required participants to listen to a sentence, identify whether the sentence was in the past or not, then identify the verb and discriminate the verb ending sound. Had the participants had more exposure to this sort of task during their interaction with the IPA, perhaps they could have performed better in the post-tests. A final limitation is the use of a single IPA, the Alexa app.

Based on these limitations, suggestions for future research would include replicating the study with a larger number of participants, extend the duration of the treatment, and include other IPAs such as those found in Google Home, Siri, and Cortana. Research could also benefit from the inclusion of other L2 phonological/pronunciation features, or a focus on other aspects of language learning (e.g., vocabulary, grammar), or learning in general.

Despite the mentioned limitations, the study contributes to our knowledge of learner experience and attitudes towards IPAs because they can further unfold the potentials and limitations of the technology. Thus, IPAs could be used as an alternative to a fluent speaker where learners can engage with in real-life communicative activities. As far as second language phonology/pronunciation is concerned, this study adds to the existing literature which explores the link between technology-enhanced listening (perceptual training) and speaking (output practice) on the acquisition of past tense allomorphy.

## Chapter Five: General Discussion

The use of technology in second language classroom, particularly intelligent personal assistants, can benefit both learners and teachers. Intelligent Personal Assistants' (IPAs) pedagogical importance lie in their ability to extend the reach of the classroom by providing learners with ample opportunities to practice speaking and listening skills outside the classroom, anytime-anywhere, and at the learners' own pace. Learners are then able to communicate with IPAs in a stress free, meaningful, and fun environment, which may contribute to an increase in the learners' willingness to communicate in the target language.

The three manuscripts featured in this dissertation shed some light on the use of IPAs for L2 learning. Each manuscript addressed one aspect of the general research questions: (a) What are language learners' perceptions of the use of IPAs as learning tools? (Manuscript A); (b) Can IPAs understand different language learners, and can these learners understand IPAs? (Manuscript B); and (c) Can IPAs help English language learners improve their receptive and productive skills? (Manuscript C).

In line with these questions, this chapter will briefly review each manuscript, summarize the key findings, and provide a discussion of the pedagogical implications of using IPAs for second language teaching and learning. It will also discuss the limitations of the dissertation, directions for future research, and concluding remarks.

### Overview, Summary, and Pedagogical Implications of the Manuscripts

This dissertation followed Cardoso's (2022) chronology for Computer Assisted Language Learning (CALL) research to explore the pedagogical affordances of IPAs. Since IPAs have already been developed, this dissertation focused on the subsequent stages of CALL development. Specifically, it examined stages 2 to 4 by exploring IPAs as tools for L2 pedagogy,

with each chapter dedicated to one (in this dissertation, affordance is defined as what a user can do with an object based on the user's capabilities). of these stages. While chapters 2 and 3 focused on examining learners' perceptions of IPAs and the tool's appropriateness as pedagogical tools, chapter 4 addressed stage 4 by examining IPAs' potential to promote learning.

***Manuscript A- Chapter 2: Practicing English with a Speaking Robot: Learners' Perceptions***

This feasibility (case) study was the first to explore IPAs' pedagogical affordance and assess its suitability for second language learning; therefore, it aimed to address this gap in the literature by examining second language learners' perceptions of a smart speaker, Amazon Echo, and its associated app, the cloud-based voice service Alexa, as a pedagogical tool. It was guided by the following research question: What are English L2 learners' perceptions of using a speaking robot (Amazon *Echo*) as a pedagogical tool?

Informed by Self Access Language Learning (SALL), which focuses on promoting learners' autonomy and independence so that learners manage their own learning (Gardner & Miller, 1999) outside the language classroom in a stress-free environment, the study investigated several variables such as Echo's ease of use, options for learner self-regulation, learner motivation and, more importantly, opportunities for learner input and output practice, i.e., to provide speaking/listening practice opportunities outside the classroom.

Seven English second language learners (three female adult participants and four female adolescent) were asked to interact with Amazon Echo. The results from an interview and survey revealed positive results for the analyzed variables: ease of use, options for self-regulation, motivation and opportunities for input and output practice. For example, the participants enjoyed using Echo and found it very user-friendly. They felt that Echo helped them acquire certain pronunciation features and vocabulary. The participants expressed that Echo provided them with

many opportunities for input and output practice and, overall, they explained that using Echo was motivating. Nevertheless, the findings also revealed some of Echo's weaknesses. For example, one main limitation is that Echo did not seem to adapt well to younger voices and to leaners with low level of English proficiency. This limitation guided the next chapter (Chapter 3) and led to further exploration of the tool, thus serving as a starting point to better understand IPAs as pedagogical tools in a learning environment that might be able to cater to learners, extend the reach of the classroom, and promote self-regulated learning.

***Manuscript B- Chapter 3: Intelligent personal assistants: Can they understand and be understood by accented L2 learners?***

The goal of this study was to examine Echo and its voice-controlled IPA, Alexa, in its ability to understand accented speech and be understood by English as a second language (ESL) learners. Specifically, it aimed to address a limitation observed in a previous study (manuscript A), which showed that accented English learners experienced difficulties understanding and being understood by the IPA. Consequently, the focus was to determine whether Alexa can communicate effectively with accented speakers (e.g., without resulting in a communication breakdown due to their fluency, foreign accent, or other speech-related issues). Emphasis was placed on how Alexa handled second language accented learners' speech in terms of accuracy, the number and types of communication breakdowns observed. In case of any interaction problems, the study also examined how the learners behaved to solve the interaction problems that they experienced with the IPA.

The study was guided by four research questions: (1) Can Alexa understand accented speech of English learners from different first languages? (2) Can the same L2 learners understand the synthesized (and consequently also "accented") speech of Alexa? (3) How do

Alexa and native-speaker judges (raters) compare in their ability to understand accented speech? And (4) When Alexa-learner communication fails, what strategies do the learners use to solve the communication breakdown?

Eleven accented English second language learners from different language backgrounds were recruited to interact with Alexa. The findings related to the first and second research questions suggested that the adopted smart speaker and its IPA, Alexa, can easily understand accented speech from English learners of different first languages, and that Alexa was able to understand accented speech at a 23% word error rate. This is considered a positive outcome from a pedagogical point of view because it shows that Alexa can accommodate to learners' pronunciation and/or lexis, without relying on strategies that do not require input processing, such as guessing from linguistic context or other non-verbal cues, as humans do when interacting with each other, or as the judges did in this study. As for the third research question, the results for transcription accuracy indicated that Alexa and the raters understand L2 accented speech with relatively similar accuracy, above 80%. Finally, the results of the fourth research question indicated that learners used a variety of strategies to mitigate the communication breakdown they experienced with Alexa. The most prominent strategies, used in the following order, consisted of repeating their questions, rephrasing them, and abandoning them. In sum, these results have shown that intelligent personal assistants such as Amazon Echo and its synthesized voice, Alexa, can understand accented L2 speech and be understood by the same accented L2 learners, with a performance relatively similar to that of two human judges. In addition, it exposed learners to oral input that is abundant and of good quality and provided them with ample opportunities for practice (both input/listening and output/speaking) through human-machine interactions.

Based on these findings, it can be concluded that IPAs can be used as teaching assistants that can interact with students and motivate them to learn on their own, at their own pace, with great potential to extend the reach of the second language classroom and consequently encourage practice and anytime-anywhere learning. However, it is still not clear to what extent IPAs can contribute to learning. Hence, following the chronology for examining CALL tools described in Cardoso (2022), the next chapter investigates whether IPAs can contribute to learning.

***Manuscript C- Chapter 4: Intelligent Personal Assistants in L2 Education: Focus on English Past -ed***

This experimental study aimed to explore the use of IPA as a tool to promote learning focusing on the acquisition of simple past tense- ed. It investigated whether Alexa App assists English learners in improving and/or developing their phonological awareness and, consequently, their perception and production of the allomorphy that characterizes the regular past tense -ed marking in English.

This study employed in its design the first four stages from the framework for pronunciation instruction that were set forth by Celce-Murcia et al (2010). According to this framework, for successful L2 pronunciation pedagogy, learners should be instructed in a way in which their phonological acquisition starts with awareness raising (stage 1), and proceeds via the development of perception (discrimination abilities- stage 2), controlled (stage 3) and guided (stage 4) oral production, towards a more spontaneous and automatized use of the target feature (stage 5). Thus, the study examined the participants' awareness to the morphophonemic forms involved in English past -ed marking, their aural perception (or ability to discriminate among the three allomorphs), and oral production. In addition, it also examined the participants' attitudes

toward the pedagogical use of IPAs in their learning. The study was thus guided by the following research questions:

1. Will the pedagogical use of a commercial IPA (Alexa) help English learners acquire English past-ed allomorphy in terms of:
    d) Phonological/sound awareness (Stage 1)
    e) Perception (or phonemic discrimination; Stage 2)
    f) Production (Stages 3 and 4)
2. What are learners' attitudes towards the pedagogical use of the Alexa App for the learning of the target English morphophonemics?

Eighteen ESL students from different language backgrounds were divided into two groups: the Alexa and the non-Alexa group. The study consisted of five main phases and five pre- and post-tests following Celce-Murcia et al.'s (2010) stages of phonological development: (1) a phonological awareness test based on a single survey that asked the participants to answer questions about their knowledge about past -ed pronunciation, using survey completion and think-aloud protocols; (2) two perception tests: while one assessed the participants ability to discriminate the three allomorphs in sentences, the other assessed the target allomorphs in words produced in isolation; and (3) two oral production tests: a read-aloud word list task for controlled production, and a role-playing for guided speech that consisted of a controlled yes/no questions task.

To answer the first research question, the participants' development across the three levels of testing was examined: awareness, perception, and production. The quantitative results of the survey for the phonological awareness test revealed that participants improved (between the pre- and post-test for the first test although this improvement was not statistically significant.

The results for the two perception tests revealed that there were no significant differences between the pre- and post-test for all measures. Similarly, the results of the production test revealed that there were no significant differences between the pre- and post-test for the read-aloud word list produced and the controlled yes/no questions.

Regarding the second question, the results indicates that Alexa has potential as a learning tool, as it had high usability scores, it motivated the participants to learn and explore the language since participants felt more confident in recognizing and distinguishing the different sounds of the past tense markings /id/ (as -ed in hunted); /d/ (hugged); /t/ (asked) and they understood that words like *walked*, *lived* and *invited* were pronounced differently and sounded differently. Also, the results showed that Alexa sparked the learners' willingness to continue to use it in their future language learning endeavours. The learners explained that the IPA was a great tool for use outside the language classroom, as a conversational partner.

This study contributes to the CALL literature by demonstrating that IPAs are valuable pedagogical tools that can extend the reach of the classroom by allowing language learners to autonomously improve aspects of their second language phonological development (e.g., awareness of past tense marking /id/, /t/, /d/). The study also contributes to our knowledge of learner experience and attitudes towards IPAs where IPAs could be used as an alternative to a fluent speaker where learners can engage with in real-life communicative activities. Finally, as far as second language phonology/pronunciation is concerned, this study adds to the existing literature which explores the link between listening (perceptual) training and output practice on the acquisition of past tense allomorphy.

## Main Takeaway Messages

The three studies mentioned above have shown that IPAs have pedagogical potential when used as part of the learning process and as an extension to the language classroom for the following reasons: The first manuscript revealed that IPAs are perceived to be enjoyable and easy to use, helpful in learning new vocabulary and certain pronunciation features, provide numerous opportunities for listening and speaking practice and, finally, they are motivating to use. The second manuscript showed that the adopted IPA (Alexa) can easily understand accented speech from English learners of different first languages and likewise, students did not have problems understanding the IPA. This study also demonstrated that learners used a variety of strategies to mitigate the communication breakdown they experienced with Alexa. These strategies consisted mostly of repeating questions, rephrasing them, and abandoning them. Finally, the third manuscript indicated that Alexa could help learners improve their phonological awareness of second language target features such as past tense marking (/id/, /t/, /d/). In addition, the study revealed that Alexa can be used as an alternative to humans or fluent speakers since the results of the Alexa group were as good as those of the non-Alexa group (if not better) which suggests that Alexa is as good as its alternative.

In sum, this dissertation has shown that IPAs can be used for language learning purposes, as they fulfil the seven criteria for CALL materials development set forth by Chapelle (2001):

1. Reliability and learner fit: IPAs are fun and easy to use by the learners and they motivate learners to practice speaking and listening in the target language, focusing on L2 features that they need.
2. Authenticity and generalizability: Despite the small number of participants presented in all three studies, the results of the studies could be generalized to other context

especially those of informal L2 learning context where learning is more spontaneous and natural.

3. Construct validity and operationalization of learning conditions: the pedagogy involved with IPAs can be motivated by many theoretical constructs such as those of Gardner & Miller's Self Access Language Learning (1999) and Long 's Interactionist Approach (1996), to cite a few.
4. Language learning potential and operationalization of learning conditions: IPAs foster repetition which is essential in learning pronunciation (Nation & Newton, 2008). IPAs also promote increased input and output practice in the form of listening and speaking in the target language.
5. Interactiveness and meaning focus: IPAs foster human-machine interactions that are meaning focused for learners.
6. Positive impact: IPAs foster learner autonomy and independence under self regulated learning where learners can understand and manage their own learning and learning environment.
7. Practicality: IPAs foster in general the notion of learning in the wild as expressed by Sauro and Zourou (2019), reflecting the dynamic and unpredictable character of the technology and learning that is often user-driven and self-and-group initiated outside the formal classroom.

## Limitations and Future Research Directions

### Limitations

Given that the manuscripts covered in this dissertation are exploratory in nature, especially manuscripts A and B, there are five limitations to be considered. The first limitation deals with the number of participants, which was low in all three studies. The low number of

participants makes it difficult to make any generalizable claims about the findings and users' attitudes towards the use of the proposed technology in general. The second limitation is the use of one IPA: Amazon Echo and its associated App Alexa. Using a variety of IPAs could have provided a better understanding of the full potential of this technology in terms of machine performance. The third limitation is privacy issues related to the adopted IPA, Alexa. Since Alexa is a cloud-based service, it is connected to massive servers where it tends to store user data. This information is gathered from users' interactions with the IPA (Lynskey, 2019). Thus, all user information is stored, and this could invade user privacy since Alexa was found to be eavesdropping and recording private conversations (Fowler, 2019), or it could be used to market product and services or even for phishing attempts. Nevertheless, Amazon is increasing its efforts to newly implement policies and practices to help mitigate some of these concerns. For example, users can now choose how long to store their voice recordings or elect not to have them saved at all. Another privacy concern deals with the activation of skills in the IPA, that is when a user allows the IPA to invoke a specific skill that can fulfill the user's request. These skills create publicized potential security and privacy concerns because of their misleading privacy policies. Because these skills are created by third parties, these companies can still change their program codes even after receiving Amazon approval, which could lead to potential leakage of sensitive personal information of the user (Lentzsch et al., 2021). The fourth limitation is the novelty effect (e.g., Cardoso, 2011). Most of the learners' positive perceptions in all three studies could have been due to the novelty effect of the technology, which may have increased the overall interest and motivation of the learners. Therefore, it would not be possible to know whether the same results would be obtained after these effects had worn out (e.g., after an extensive

experience with IPAs as pedagogical tools). To achieve that, the duration of the studies, intervention or learning activity should be extended, leading us to the fifth and final limitation.

The final limitation is the short duration of the intervention and learning activities presented in all three studies. The participants were only given between 30 to 45 minutes of time to interact with the IPA per session for a total of 4.5h ; had they been provided with a longer period for interaction and practice, it is possible that they could have developed some other aspects of phonological knowledge, such as perception and production.

### Future Directions

The future directions discussed here aim to advance research in CALL in relation to the use of IPAs. One possible future direction is to consider a larger selection of IPAs and to compare the affordances of these IPAs in general and their individual limitations, and potentially suggest the ones that are more suitable for second language pedagogy. Another idea is to further examine interactions between IPAs and learners. The studies mentioned above only looked at individual interactions between a learner and an IPA. It would be interesting to look at a combination of interactions between a group of second language learners and an IPA and inspect the strategies that they use when collaborating with each other while they communicate with the IPA. In addition, one could investigate the interaction with IPAs in the context of the acquisition of language skills informally as in the case of "learning in the wild". Another aspect worthy of research is to investigate teachers' perceptions toward the use of IPAs for language teaching purposes. Finally, another important direction for research would be to delve into whether the extended use of IPAs may contribute to the learning of other language features such as oral fluency and prosody (e.g., intonation for questions, rhythm).

## Concluding Remarks

The goal of this dissertation was to investigate a popular intelligent personal assistant Alexa, the voice of Amazon's smart speaker Echo, as a pedagogical tool in second language learning and as a supportive tool for teachers and learners. This dissertation examined Alexa through stages that started with the exploration of the tools, then proceeded with the examination of their pedagogical suitability, and culminated with the assessment of their pedagogical effectiveness. In this dissertation, I was able to demonstrate that IPAs have the potential to extend the reach of the classroom, promote learner-machine interaction, and enhance and aid students' language learning experience. While future research is still fully needed to understand the importance of IPAs in second language pedagogy, this dissertation has paved the way and offered some evidence on how pedagogically effective this technology can be.

## References


Alastuey, M. C. (2012). Focus on form and negotiation of meaning in synchronous voice-based computer mediated communication: Effect of dyad. *Procedia - Social and Behavioral Sciences*, *34*, 39-44.

Alipanahi, F., & Mahmoodi, R. (2015). Corrective feedback via e-mail on the correct use of past tense among Iranian EFL learners. South African Journal of Education, 35(4), 1-19. doi:10.15700/saje.v35n4a1159alip

AL-Jarf, R. S. (2005). The effects of online grammar instruction on low proficiency EFL college students' achievement. *Asian EFL Journal 7*(4), 166-190. Available at: https://www.academia.edu/2358840/The_effects_of_online_grammar_instruction_on_low_proficiency_EFL_college_students_achievement.

Baleghizadeh, S., & Oladrostam, E. (2010). The effect of mobile assisted language learning (MALL) on grammatical accuracy of EFL students. *MEXTESOL Journal, 34*(2), 1–10

Beatty, K. (2003). *Teaching and Researching Computer-Assisted Language Learning* (Applied Linguistics in Action) (illustrée ed.). Longman.

Blake, R. (2000). Computer mediated communication: A window on L2 Spanish interlanguage. *Language Learning & Technology, 4*, 120–136.

Bowles, M. A. (2010). *The think-aloud controversy in second language research*. Routledge.

Braul, B. (2006). ESL teacher perceptions and attitudes toward using computer-assisted language learning (CALL): Recommendations for effective CALL practice, (MA Dissertation), Department of Secondary Education, Edmonton, Alberta.

Canale, M., & Swain,M. (1980). Theoretical basis of communicative approaches to second language teaching and resting. *Applied Linguistics,1*(1), 1-47.

Cardoso, W. (2011). Learning a foreign language with a learner response system: The students' perspective. *Computer Assisted Language Learning, 24*(5)*,* 393-417.

Cardoso, W. (2018). Learning L2 pronunciation with a text-to-speech synthesizer. In Peppi Taalas, Linda Bradley, and Sylvie Thouësny (Eds.), *Language Learning as Exploration and Encounters*; Selected papers from EuroCALL (pp. 16-21). Dublin: Research-publishing.net.

Cardoso, W. (2022). Technology for Speaking Development. In T. Derwing, M. Munro, & R. Thomson (Eds), *Routledge Handbook on Second Language Acquisition and Speaking*. Routledge, Taylor & Francis Group.

Celce-Murcia, M., Brinton, D.M., Goodwin, J.M. & Griner, B. (2010). *Teaching pronunciation: A course book and reference guide*. 2nd edition. Cambridge, UK: Cambridge University Press.

Chapelle, C. (2001). *Computer applications in second language acquisition: Foundations for teaching, testing and research*. Cambridge: Cambridge University Press.

Chapelle, C. (2003). *English language learning and technology: Lectures on applied linguistics in the age of information and communication technology*. Amsterdam/Philadelphia: John Benjamins.

Chapelle, C. (2008). Computer Assisted Language Learning. In B. Spolsky Editor and F.M. Hult Editor (Eds.),*The handbook of educational linguistics* (pp.585-595). Blackwell Publishing Ltd. https://doi.org/10.1002/9780470694138.ch41

Clark, R. E., & Sugrue, B. M. (1991). Research on instructional media. In G. Anglin (Ed.),

*Instructional technology: Past, present and future*. Englewood, Colorado, U.S.A.: Libraries Unlimited, pp. 327-343

Clarke, T.B. & Nelson, C. (2012). Classroom community, pedagogical effectiveness, and learning outcomes associated with Twitter use in undergraduate marketing courses. *Journal for Advancement of Marketing Education. 20*(2). 29-38.

Collins, L., Trofimovich, P., White, J., Cardoso, W., & Horst, M. (2009). Some input on the Easy/Difficult grammar question: An empirical study. *The Modern Language Journal, 93*, 336-353.

Collins, L., & Muñoz, C. (2016). The foreign language classroom: Current perspectives and future considerations. *The Modern Language Journal, 100(*1), 133-147.

Coniam, D. (1999). Voice recognition software accuracy with second language speakers of English, *System, 27*(1), 49–64.

Cotterall, S. & Reinders, H. (2001). Language learners learning independently: How autonomous are they? Toegepaste Taalwetenschap in Artikelen, 65, 85-96. doi:10.1075/ttwia.65.09rei

De Barcelos Silva, A., Gomes, M. M., da Costa, C. A., da Rosa Righi, R., Barbosa, J. L. V., Pessin, G., ... & Federizzi, G. (2020). Intelligent personal assistants: A systematic literature review. *Expert Systems with Applications*, 147, 113193.

DeKeyser, R. M. (2005). What makes learning second-language grammar difficult? A review of issues. *Language Learning, 55*, 1-25.

Delatorre, F. (2010). The role of orthography on the production of regular verbs ending in ed by Brazilian EFL learners. *In Proceedings of the 9th Seminário do Círculo de Estudos Linguísticos do Sul* (pp. 1-13). Florianópolis: Federal University of Santa Catarina.

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition, 19*(1), 1-16.

Derwing, T. M., Munro, M. J., & Carbonaro, M. (2000). Does popular speech recognition software work with ESL speech? TESOL Quarterly, 34(3), 592-603.

Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching, 42*, 276–490.

De Vries, B. P., 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

Diemer, T. T., Fernandez, E., & Streepey, J. W. (2013). Student perceptions of classroom engagement and learning using iPads. *Journal of Teaching and Learning with Technology*, *1*(2), 13-25.

Dizon, G. (2017). Using intelligent personal assistants for second language learning: A case study of Alexa. *TESOL Journal, 8*(4), 811-830.

Dörnyei, Z. (2009). *The psychology of second language acquisition*. Oxford: Oxford University Press.

Driscoll, M. & Carliner, S. (2005.) *Advanced web-based training strategies: Unlocking instructionally sound online learning*. Wiley/Pfeiffer.

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition, 24*(2), 143–188. http://www.jstor.org/stable/44486609

Enge, E. (2017). 31 must know stats about the continuing growth of voice commands with smartphones [blog post]. Retrieved from: https://www.stonetemple.com/mobile-voice-usage-trends/

Felix, U. (2001). The web's potential for language learning: the student's perspective. *ReCall, 13*(1), 47-58.

Fernandez-Garcia, M., & Martinez Arbelaiz, A. (2002). Negotiation of meaning in nonnative speaker-nonnative speaker synchronous discussions. *Computer Assisted Language Instruction Consortium (CALICO), 19*(2), 279-294.

Fernandez-Garcia, M., & Martinez Arbelaiz, A. (2003). Learners' interactions: a comparison of oral and computer-assisted written conversations. *ReCALL, 15*, 113–136

Fiori, M. L. (2013). The development of grammatical competence through synchronous computer-mediated communication. *Computer Assisted Language Instruction Consortium (CALICO), 22*(3), 567-602. doi:10.1558/cj.v22i3.567-602

Fowler, G. A. (2019, May 6). Alexa has been eavesdropping on you this whole time. *The Washington Post*. https://www.washingtonpost.com/technology/2019/05/06/alexa-has-been-eavesdropping-you-this-whole-time/

Friggard, A. (2002). *Does the computer lab improve student performance on vocabulary, grammar, and listening comprehension?* (Master's thesis).

Gagne, R. (1985). The Conditions of Learning (4th ed.). New York: Holt, Rinehart & Winston

Gardner, D., & Miller, L. (1999). Establishing self-access: From theory to practice. Cambridge, UK: Cambridge University Press.

Gardner, R. (1985). Social psychology and second language learning: The role of attitudes and motivation. Edward Arnold.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston:Houghton Mifflin.

Graus, J. & Coppen, P.-A. (2015). Defining grammatical difficulty: A student teacher perspective. *Language Awareness, 24*(2), 101-122.

Gregersen, T. & MacIntyre, P. D. (2014). *Capitalizing on language learners' individuality: From premise to practice*. Bristol: Multilingual Matters.

Godwin-Jones, R. (2009). Emerging technologies focusing on form: Tools and strategies. *Language Learning & Technology, 13*(1), 5-12.

Goksel-Canbek, N., & Mutlu, M. E. (2016). On the track of Artificial Intelligence: Learning with Intelligent Personal Assistants. *International Journal of Human Sciences*, *13*(1), 592.

Handley, Z., & Hamel, M-J. (2005). Establishing a methodology for benchmarking speech synthesis for computer-assisted language learning (CALL). *Language Learning & Technology, 9*(3), 99–120. http://dx.doi.org/10125/44034

Heift, T., & Vyatkina, N. (2017). Technologies for Teaching and Learning L2 Grammar. The Handbook of Technology and Second Language Teaching and Learning, 26-44. doi:10.1002/9781118914069.ch3

Holland, V. M., Kaplan, J. D., & Sabol, M. A. (1999). Preliminary tests of language learning in a speech interactive Graphics Microworld. *Computer Assisted Language Instruction Consortium (CALICO), 16*(3), 339-359.

Hsu, L. (2015). An empirical examination of EFL learners' perceptual learning styles and acceptance: SR-based computer-assisted pronunciation training. *Computer Assisted Language Learning, 28*, 1-20.

Jepson, K. (2005). Conversations—and negotiated interaction—in text and voice chat rooms. *Language Learning & Technology, 9*, 79–98.

Jia, G. & Fuse, A. (2007). Acquisition of English grammatical morphology by native Mandarin-speaking children and adolescents: Age-related differences. *Journal of Speech, Language and Hearing Research, 50*, 1280-1299.

Jiang, N., Novokshanova, E., Masuda, K. & Wang, X. (2011). Morphological congruency and the acquisition of L2 morphemes. *Language Learning, 61*(3), 940-967.

Johnson, L., Adams Becker, S., Estrada, V. & Freeman, A. (2014). *NMC Horizon Report: 2014 K-12 Edition. Austin, Texas: The New Media Consortium*. Retrieved September 25, 2018 from https://www.learntechlib.org/p/147472/.

Khademi, H. & Cardoso, W. (in press). Learning beyond the walls of classrooms: Exploring Google Translate and its speech capabilities for second language pronunciation learning. *Computer Assisted Language Learning.*

Kim, H., Kwon, Y. (2012). Exploring smartphone applications for effective mobile-assisted language learning. *Multimedia-Assisted Language Learning,15*(1), 31-57

Kinshuk, & Graf, S. (2012) Ubiquitous learning. In Norbert M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 3361-3363).

Klein, E., Stoyneshka, I., Adams, K., Rose, T., Pugash, Y., & Solt, S. (2004). Past tense affixation in L2 English: The effects of lexical aspect and perceptual salience. In *BUCLD 28: Proceedings of the 28th Annual Boston University Conference on*

*Language Development: Online supplement. Available from www. bu. edu/bucld/proceedings/supplement*.

Kukulska-Hulme. A. (2012). Language learning defined by time and place: A framework for next generation designs. In J. E. Díaz-Vera (Ed.), *Left to my own devices: Learner autonomy and mobile assisted language learning* (pp. 1–13). Bingley, UK: Emerald Group Publishing Limited.

Lah, U., Lewis, J. R., & Šumak, B. (2020). Perceived usability and the modified technology acceptance model. *International Journal of Human–Computer Interaction*, *36*(13), 1216–1230. https://doi.org/10.1080/10447318.2020.1727262

Lam, Y., & Lawrence, G. (2002). Teacher-student role redefinition during a computer-based second language project: Are computers catalysts for empowering change? *Computer Assisted Language Learning, 1*5(3), pp. 295-315.

Lawrence, B. M. (2016). iPad acceptance by English learners in Saudi Arabia. *English Language Teaching*, *9*(12), 34-46.

Lee, L. (2001). Online interaction: negotiation of meaning and strategies used among learners of Spanish. *ReCALL, 13*(2), 232–244.

Lee, J.Y., (2009). *The effect of computer-mediated communication (CMC) interaction on L2 vocabulary acquisition: A comparison study of CMC interaction and face-to-face interaction*. (Master's thesis, Iowa State University, U.S.A.).

Lentzsch, L., Shah, S. J., Benjamin Andow, B., Degeling, M., Das, A., & Enck, W. (2021, February). *Hey Alexa, is this skill safe?: Taking a closer look at the alexa skill*

*ecosystem* . Paper presented at the Network and Distributed Systems Security (NDSS) Symposium, San Diego, CA, USA. https://dx.doi.org/10.14722/ndss.2021.23111 Retrieved October 20, 2021, from https://anupamdas.org/paper/NDSS2021.pdf.

Liakin, D., Cardoso, W., Liakina, N. (2015). Learning L2 pronunciation with a mobile speech recognizer: French /y/. *Computer-Assisted Language Instruction Consortium (CALICO), 32*(1), 1-25.

Lightbown, P. M. (2000). Classroom SLA research and second language teaching. *Applied Linguistics, 21*, 431-462.

Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie & T. Bhatia (Eds), *Handbook of Second Language Acquisition* (pp. 413–468). San Diego, CA: Academic Press.

Lynskey , D. (2019, October 9). Alexa, are you invading my privacy?' – the dark side of our voice assistants. *The Guardian.* https://www.theguardian.com/technology/2019/oct/09/alexa-are-you-invading-my-privacy-the-dark-side-of-our-voice-assistants

Mansor, N. (2007). Collaborative learning via email discussion: Strategies for ESL writing classroom. *The Internet ESL Journal, 13*(3).

Mariano, M. H. (2009) The influence of training and instruction on the production of words ending in -ed by Brazilian EFL learners. 2009. 89f. (Master's thesis), Universidade Federal de Santa Catarina, Florianópolis.

McDonald, J. L. & Roussel, C. C. (2010). Past tense grammaticality judgment and production in non-native and stressed native English speakers. *Bilingualism, Language and Cognition, 13*(4), 429-448.

Mohamad, F. (2009). Internet-based grammar instruction in the ESL classroom.

*International Journal of Pedagogies and Learning, 5*(2), 34-48

Nagata, N. (1996). Computer vs. workbook instruction in second language acquisition. *Computer Assisted Language Instruction Consortium (CALICO),* 14(1), 53-75.

Nation, I.S.P. & Newton, J. (2008). *Teaching ESL/EFL listening and speaking*. Routledge.

Negi, S. J. (2009). The role of teachers' non-verbal communication in ELT classroom. *Journal of Nelta,14*(1-2), 101-110.

Neri, A., Cucchiarini, C., & Strik, W. (2003). Automatic speech recognition for second language learning: How and why it actually works. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *ICPhS 15* (pp. 1157-1160). Barcelona, Spain: ICPhS.

Nunan, D., & Richards, J. C. (2014). *Language learning beyond the classroom*. Routledge

Nutta, J. (1998). Is computer-based grammar instruction as effective as teacher directed grammar instruction for teaching L2 structures? *Computer Assisted Language Instruction Consortium (CALICO),, 16*(1), 4962.

Ozkan, D. & Morency, L.P. (2013). *Prediction of visual backchannels in the absence of visual*

*context using mutual influence.* In R. Aylett, B. Krenn, C. Pelachaud & H. Shimodaira

(eds.), International Conference on Intelligent Virtual Humans (pp. 189-202): Springer.

Pellettieri, J. (2000). Negotiation in cyberspace: The role of chatting in the development of grammatical competence. In M. Warschauer & R. Kern (Eds.), Network-based language teaching: Concepts and practice (pp. 59-86). New York: Cambridge University Press.

Peters, M., Weinberg, A., & Sarma, N. (2009). To like or not to like! Student perceptions of technological activities for learning French as a second language at five Canadian universities. *Canadian Modern Language Review*, *65*(5), 869-896. doi:10.3138/cmlr.65.5.869

Pirasteh, P. (2014). The effectiveness of Computer-assisted Language Learning (CALL) on learning grammar by Iranian EFL Learners. Procedia - Social and Behavioral Sciences, 98, 1422-1427. doi:10.1016/j.sbspro.2014.03.561

Rao, L. (2017, May). Amazon is dominating the voice-assisted speaker market. *Fortune.* Retrieved from http://fortune.com/2017/05/08/amazon-echo-alexa-speakers

Rifkin, B. (2005). A ceiling effect in traditional classroom foreign language instruction: Data from Russian. *The Modern Language Journal, 89*(1), 3-18

Ripley, A. (2008). *Using Technology to Develop Early Phonological Awareness Skills*. (Master's Theses, State University of New York College, Brockport).

Rogers, C. L., Dalby, J. M., & DeVane, G. (1994). Intelligibility training for foreign-accented speech: A preliminary study. *Journal of the Acoustical Society of America*, *96*(5), 33-48.

Rouhshad, A., Wigglesworth, G., & Storch, N. (2015). The nature of negotiations in face-to-face versus computer-mediated communication in pair interactions. *Language Teaching Research*, *20*(4), 514-534.

Salaberry, M. R. (2000). L2 morphosyntactic development in text based computer mediated communication. Computer Assisted Language Learning, 13(1), 5-27.

Saldaña, J. (2009). *The coding manual for qualitative researchers*. Los Angeles: Sage.

Shaalan 1, K. F. (2005). An intelligent computer assisted language learning system for Arabic learners. *Computer Assisted Language Learning 18* (1-2), 81-109.

Sauro, S., & Zourou, K. (2019). What are the digital wilds? *Language Learning & Technology*, *23*(1), 1–7. https://doi.org/10125/44666

Segan S. (2018, October). The best smart speakers of 2018. *PCMag*. Retrieved from https://www.pcmag.com/article/357520/the-best-smart-speakers

Segers, E., & Verhoeven, L. (February, 2005). Long-term effects of computer training of phonological awareness in kindergarten. *Journal of Computer Assisted Learning, 2(*1), 17-27. Retrieved October 24, 2007, from Academic Search Premier database

Smith, B. (2003). Computer-mediated negotiated interaction: an expanded model. *The Modern Language Journal, 87*, 38–57.

Solt, S., Pugach, Y., Klein, E. C., Adams, K., Stoyneshka, I., & Rose, T. (2003). L2 perception and production of the English regular past: Evidence of phonological effects. *In Annual Boston University Conference on Language Development*. *Boston, MA.*

Sotillo, S. (2000). Discourse functions and syntactic complexity in synchronous and asynchronous communication. *Language Learning & Technology, 4*(1), 82-119. Retrieved April 26, 2005, from http://llt.msu.edu/vol4num1/sotillo/default.html

Tarone, E. (1980). Communication strategies, foreigner talk, and repair in interlanguage. *Language Learning*, *30*(2), 417-428.

Taylor, R., & Gitsaki, C. (2004). Teaching WELL and loving IT. In S. Fotos, & C. Browne (Eds.), New perspectives on CALL for second language classrooms (pp. 131-147). Mahwah, New Jersey, U.S.A.: Lawrence Erlbaum Associates.

Thornton, P. & Houser, C. (2005). Using mobile phones in English education in Japan. *Journal of Computer Assisted Learning, 21,* 217–228.

Timucin, M. (2006), Implementing CALL in the EFL context. *ELT Journal*, *60*(3), pp. 262-271.

Varonis, E., & Gass, S. (1985). Non-native/non-native conversations: A model for the negotiation of meaning. *Applied Linguistics, 6*, 71-90.

Wang, S., & Smith, S. Reading and grammar learning through mobile phones. *Language Learning & Technology, 17*(3), 117–134. Retrieved from http://llt.msu.edu/issues/october2013/wangsmith.pdf

Watson-Gegeo, K. A. (1988). Ethnography in ESL: Defining the essentials. TESOL Quarterly, 22(4), 575-592.

Weizenbaum, J. (1966). ELIZA – a computer program for the study of natural language communication between man and machine. *Computational Linguistics, 9*(1), 36-45.

Wong, W. (2005). *Input enhancement: From theory and research to the classroom*. New York, NY: McGraw Hill.

Yanguas, Í. (2010). Oral computer-mediated interaction between L2 learners: It's about time. Language Learning & Technology, 14, 72–93.

Zhao, Y. (2003). Recent developments in technology and language learning: A literature review and meta-analysis. *Computer Assisted Language Instruction Consortium (CALICO),* 21(1), 7-27.

# Appendices

## Appendix A: Consent Form

### CONSENT TO PARTICIPATE IN RESEARCH ABOUT USING A SPEAKING DEVICE (ECHO)

I understand that I have been asked to participate in a program of research being conducted by *Souheila Moussalli* and *Dr. Walcir Cardoso* of *the Department of Education* of Concordia University. walcir@education.concordia.ca

A. PURPOSE

I have been informed that the purpose of the research is to investigate learner's perceptions of Amazon Echo, a wireless speaker and voice command device with a voice recognizer Alexa that has the potential to be used for oral English practice and improve learners' input and output practice.

B. PROCEDURES

I understand that I will participate in one session that will last approximately one hour. The practice will include a series of short oral interactions with Echo, which will be audio-recorded.

The researchers will know my name, but my name will not be revealed to anyone else. No teacher, student, or school names will be used to identify any digital or electronic files. Codes and pseudonyms will be assigned.

C. RISKS AND BENEFITS

I understand that the researcher is interested in examining my general perceptions of Amazon Echo via an oral interview at the end of the session. I understand that the researchers will not evaluate me individually, but will aggregate the data across all the participants in the activity.

D. CONDITIONS OF PARTICIPATION

- I understand that I am free to withdraw my consent and discontinue my participation at anytime without negative consequences.

- I understand that my participation in this study is confidential (i.e., the researcher will know, but will not disclose my name)

- I understand that the data from this study may be published and that I may receive a copy of the final research report when the study has been completed (expected to take several months) by contacting Ms. Souheila Moussalli and Dr. Walcir Cardoso

I HAVE CAREFULLY READ THE ABOVE AND UNDERSTAND THIS AGREEMENT. I FREELY CONSENT AND VOLUNTARILY AGREE TO PARTICIPATE IN THIS STUDY.

NAME (please print) _____

SIGNATURE _____

If at any time you have questions about the proposed research, please contact *Ms. Souheila Moussalli* and *Dr. Walcir Cardoso*, at the email address above. If at any time you have questions about your rights as a research participant, please contact the Research Ethics and Compliance Advisor, Concordia University, 514.848.2424 ex. 7481 ethics@alcor.concordia.ca

## Appendix B: Consent Form

### CONSENT TO PARTICIPATE IN RESEARCH ABOUT USING A SPEAKING DEVICE (ECHO)

I _____ consent for my child _____ to participate in a program of research being conducted by Souheila Moussalli and Dr. Walcir Cardoso of the Department of Education of Concordia University.
walcir@education.concordia.ca

A. PURPOSE

I have been informed that the purpose of the research is to investigate learner's perceptions of Amazon Echo, a wireless speaker and voice command device with a voice recognizer Alexa that

has the potential to be used for oral English practice and improve learners' input and output practice.

## B. PROCEDURES

I understand that my child will participate in one session that will last approximately one hour. The practice will include a series of short oral interactions with Echo, which will be audio-recorded.

The researchers will know my child's name, but my child's name will not be revealed to anyone else. No teacher, student, or school names will be used to identify any digital or electronic files. Codes and pseudonyms will be assigned.

## C. RISKS AND BENEFITS

I understand that the researcher is interested in examining my child's general perceptions of Amazon Echo via an oral interview at the end of the session. I understand that the researchers will not evaluate my child individually, but will aggregate the data across all the participants in the activity.

## D. CONDITIONS OF PARTICIPATION

- I understand that I am free to withdraw my child's consent and discontinue my chidl's participation at anytime without negative consequences.

- I understand that my child's participation in this study is confidential (i.e., the researcher will know, but will not disclose my name)

- I understand that the data from this study may be published and that I may receive a copy of the final research report when the study has been completed (expected to take several months) by contacting Ms. Souheila Moussalli and Dr. Walcir Cardoso

I HAVE CAREFULLY READ THE ABOVE AND UNDERSTAND THIS AGREEMENT. I FREELY CONSENT AND VOLUNTARILY AGREE FOR MY CHILD TO VOLUNTARILY PARTICIPATE IN THIS STUDY.

CHILD'S NAME (please print)

_____

PARENT/GUARDIAN'S NAME (please print)

_____

PARENT/GUARDIAN'S SIGNATURE

_____

My name is _____and I am doing a research study about a speaking Robot Echo and about my opinions about it. A research study is a way to learn more about something. One of your parents or guardians has already given their permission for you to be part of this study, and now it's your turn to decide. If you decide that you want to be part of this study, you will be asked to practice speaking and listening with Echo for one hour which will be audio-recorded. When we are finished with this study, we will write a report about what was learned. This report will not include your name or that you were in the study. You do not have to be in this study if you do not want to be. If you have questions or decide to stop after we begin, that's okay; just tell your parent or guardian or Ms. Souheila Moussalli or Dr. Walcir Cardoso.

If you want to be in this study, please write your name and sign below.

Check one:

_____ I want to be in the study.

_____ I do NOT want to be in the study.

Your name: _____

Your signature: _____

If at any time you have questions about the proposed research, please contact Ms. Souheila Moussalli and Dr. Walcir Cardoso, at the email address above. If at any time you have questions about your rights as a research participant, please contact the Research Ethics and Compliance Advisor, Concordia University, 514.848.2424 ex. 7481 ethics@alcor.concordia.ca

## Appendix C: Language Background Questionnaire

Date of testing: _____ Participant Code_____

Language Background Questionnaire

Name: _____ Gender: Male____ Female _____ Other _____

Phone number: _____ Email address: _____

Is your hearing normal as far as you know? Yes: _____ No: _____

Date of birth: _____ Birthplace (City, Country): _____

**Native language**:

Your native language is:

Your mother's native language is:

Your father's native language is:

Were you exposed to this language since birth? Yes: _____ No: _____

What language do you speak at home now? _____

In what language did you attend school? Please choose the appropriate one (you may check more than one option):

**Elementary school**: ( ) English
( ) French
( ) Other (please specify):

**High school**: ( ) English
( ) French
( ) Other (please specify):

**University**: ( ) English
( ) French
( ) Other (please specify):

**Second language**:

Your second language is: 1-English 2- Other: _____

Age when you started learning your second language is: _____ years old

**Your Knowledge of English**

Please rate your ability to speak, listen to, read and write in **English** by using the scales in the box below. Please note that **1= extremely poor** and **9= extremely fluent**

| Speaking | Listening | Reading | Writing |
|---|---|---|---|
| 1 2 3 4 5 6 7 8 9 | 1 2 3 4 5 6 7 8 9 | 1 2 3 4 5 6 7 8 9 | 1 2 3 4 5 6 7 8 9 |

At what age, did you start learning English? _____ years old

Do you speak or know any other languages? YES ( ) NO ( )
If **YES**, please list them below and circle the options that correspond to WHERE you learned it and your overall PROFICIENCY.

| LANGUAGE | WHERE YOU LEARNED | PROFICIENCY (circle one) |
|---|---|---|
| | Classroom \| Independently | Beginner \| intermediate \| advanced \| native |
| | Classroom \| Independently | Beginner \| intermediate \| advanced \| native |
| | Classroom \| Independently | Beginner \| intermediate \| advanced \| native |

**Your Knowledge and Use of Technology**:

| | |
|---|---|
| Do you like using technology for general purposes? For example, to interact with friends via social media, to game, to have fun, etc. | ( ) Yes<br>( ) A little<br>( ) No (not a lot) |
| Do you use any kind of technology for learning? | ( ) Yes<br>( ) No |

If yes, do you like using technology for learning? ( ) Yes

( ) No

If yes, what do you use?

How do you use it?

## Appendix D: ECHO/ ALEXA Target questions

**Language**:

1. Alexa, spell _____
2. Alexa, ask pronunciations how to pronounce B. I. T. S.
3. Alexa, ask pronunciations to pronounce P. I. L. A. N. I.
4. Alexa, open Translated. Ask translated: 'love in French" (translates into 50 languages*)
5. Alexa, repeat after me: _____
6. Alexa, what is the definition of _____?
7. Alexa, what is the antonym of _____?
8. Alexa, tell me a story.
9. Alexa, open interactive stories.
10. Alexa, open Dr. Speech.
11. Alexa, read me a poem.
12. Alexa, say the alphabets
13. Alexa, ask TED TALKS to find talks about _____
14. Alexa, open grammar tool.

**General information: math- science- geography- history- Wikipedia-chemistry**

1. Alexa, how tall is Mount Everest?
2. Alexa, what is the capital of_____?
3. Alexa, how far is the moon?
4. Alexa, what is 2 +2?
5. Alexa, what is the square root of 64?
6. Alexa, How many cups in a litre?
7. Alexa, Wikipedia: " Niagara Falls"
8. Alexa, what is the definition of _____?
9. Alexa, ask Today in History about a famous birthday today.
10. Alexa, ask Today in History to tell me a random fact about today.

**Games: do you know any of the games?**

1. Alexa, tell me a joke.
2. Alexa, tell me a riddle.
3. Alexa, play rock paper scissors
4. Alexa, knock knock (enable skill if doesn't work*)

5. Alexa, Simon says _____
6. Alexa, play Jeopardy.

# Appendix E: Survey

## Amazon Echo – Final Survey

Please select the option that best reflects your perception of using Echo, based on your recent experience with it. Note that (**1**) indicates **"strongly DISAGREE"** while (**5**) indicates **"strongly AGREE".**

1. *Echo* **MOTIVATED** me to learn more **ENGLISH**.
   **Strongly DISAGREE** (1) (2) (3) (4) (5) **Strongly AGREE**
2. *Echo* **MOTIVATED** me to learn about **English PRONUNCIATION**.
   **Strongly DISAGREE** (1) (2) (3) (4) (5) **Strongly AGREE**
3. *Echo* **MOTIVATED** me to learn about **English VOCABULARY**.
   **Strongly DISAGREE** (1) (2) (3) (4) (5) **Strongly AGREE**
4. *Echo* **MOTIVATED** me to become more **FLUENT** in English.
   (Fluency = your ability to speak fast and clearly and be understood by others)
   **Strongly DISAGREE** (1) (2) (3) (4) (5) **Strongly AGREE**
5. I felt more comfortable **SPEAKING ENGLISH** using *Echo* than I would in **other types of classroom activities** (ex.: role-playing, listen and repeat, group work).
   **Strongly DISAGREE** (1) (2) (3) (4) (5) **Strongly AGREE**
6. I felt more comfortable **SPEAKING ENGLISH while** using *Echo* than I would **in front of the *teacher***.
   **Strongly DISAGREE** (1) (2) (3) (4) (5) **Strongly AGREE**
7. I felt more comfortable **SPEAKING ENGLISH while** using *Echo* than I would **in front of other students**.
   **Strongly DISAGREE** (1) (2) (3) (4) (5) **Strongly AGREE**
8. I learned about **English PRONUNCIATION** using *Echo*.
   **Strongly DISAGREE** (1) (2) (3) (4) (5) **Strongly AGREE**
9. Echo can help me **improve my ENGLISH**.
   **Strongly DISAGREE** (1) (2) (3) (4) (5) **Strongly AGREE**
10. Echo can help me **improve my PRONUNCIATION**.
    **Strongly DISAGREE** (1) (2) (3) (4) (5) **Strongly AGREE**
11. *Echo can* help me **improve my VOCABULARY**.
    **Strongly DISAGREE** (1) (2) (3) (4) (5) **Strongly AGREE**
12. *Echo can* help me **improve my FLUENCY**.
    (Fluency = your ability to speak fast and clearly and be understood by others)
    **Strongly DISAGREE** (1) (2) (3) (4) (5) **Strongly AGREE**
13. I would like to use *Echo* **to learn other languages**.
    **Strongly DISAGREE** (1) (2) (3) (4) (5) **Strongly AGREE**
14. *Echo* is a **great tool to learn a language**.
    **Strongly DISAGREE** (1) (2) (3) (4) (5) **Strongly AGREE**
15. Echo can **understand me**.
    **Strongly DISAGREE** (1) (2) (3) (4) (5) **Strongly AGREE**
16. I can **understand Echo.**
    **Strongly DISAGREE** (1) (2) (3) (4) (5) **Strongly AGREE**
17. Overall, I **enjoyed** using *Echo* in this project.
    **Strongly DISAGREE** (1) (2) (3) (4) (5) **Strongly AGREE**

## Appendix F: Consent Form


INFORMATION AND CONSENT FORM

Study Title: Personal robots and L2 learning: Can they understand foreign accents and be understood by learners?

Researcher: Souheila Moussalli

Contact Information: e-mail: souheilamoussally@hotmail.com

Faculty Supervisor: Dr. Walcir Cardoso

Contact Information: (514) 848-2424 ext. 2451 / e mail: walcir.cardoso@concordia.ca

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 students' perceptions of Echo, a wireless personal assistant with a built-in voice recognizer that has the potential to be used for English language practice, particularly listening and speaking skills.

B. PROCEDURES

If you participate, you will be asked to fill out a questionnaire and a survey, do some practice exercises that will include a series of short oral interactions with Echo which will be video- and

audio-recorded, and participate in a semi-structured interview which will be also video- and audio-recorded.

In total, participating in this study will take around 45 minutes to 1hour.

## C. RISKS AND BENEFITS

There are no risks in participating in this research. Potential benefits include: improving overall pronunciation and vocabulary knowledge, and exposure to a new technology for language learning that promotes learning in a stress-free environment.

## D. CONFIDENTIALITY

We will examine your general perceptions of Echo via a survey and an oral interview as part of this research. In this study,

We will not allow anyone to access the information, except people directly involved in conducting the research. We will only use the information for the purposes of the research described in this form.

The information gathered will be identifiable. That means it will have your name directly on it.

We will protect the information by storing it on a password-protected laptop computer in a secure place in the researcher's office at Concordia University (FG 6.431). The researchers will know your name, but your name will not be revealed to anyone else. No teacher, student, or school names will be used to identify any electronic files. Codes and pseudonyms will be assigned.

We intend to publish the results of the research. However, it will not be possible to identify you in the published results.

We will destroy the information five years after the end of the study.

## E. CONDITIONS OF PARTICIPATION

You do not have to participate in this research. It is purely your decision. If you 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.

There are no negative consequences for not participating, stopping in the middle, or asking us not to use your information

If you decide that you don't want us to use your information, you must tell the researcher before analysis of the study is conducted, that is before April 15, 2017.

F. 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 G: Assessment of Echo's speech (by L2 leaners)**

# Rubric for assessing Echo's speech (Participants)

Please select the option that best reflects how much you think you can understand Echo and Echo's pronunciation.

| Question | Explanation |
|---|---|
| How much do you think you can **understand** Echo | This term refers to how much effort it takes to understand what Echo is saying. If you can understand with ease, then a speaker is highly easy to understand (5 below). However, if you struggle and must listen very carefully, or in fact cannot understand what is being said at all, then a speaker is very difficult to understand (1 below). |
| How do you evaluate Echo's voice in **comparison with human** voice? | This refers to how much Echo's voice is similar to human voice. If you are not doing an effort listening and the voice is pleasant, not forced and dramatized, then it is natural (5 below). If you are making a great effort listening and the voice is not pleasant, forced and dramatized, then it is not natural (1 below). |

Global judgment of Echo's pronunciation:

1- How difficult do you think it is to **understand** Echo?

**Difficult to understand** **(1)** **(2)** **(3)** **(4)** **(5)** **Easy to understand**

2- How do you evaluate Echo's voice:

**Not natural** **(1)** **(2)** **(3)** **(4)** **(5)** **natural**

3- How much can you really understand Echo? For this question, ask Echo "Alexa, are you in love? " and write down what Echo said, based on what you heard:

## Appendix H: Assessment of Echo's speech (by judges)

# Rubric for assessing participants' speech (Raters)

Please select the option that best reflects your general perceptions about the speakers' speech and pronunciation.

| Question | Explanation |
|---|---|
| How much do you think you can **understand** the speaker? | This term refers to how much effort it takes to understand what someone is saying. If you can understand with ease, then a speaker is highly easy to understand (5 below). However, if you struggle and must listen very carefully, or in fact cannot understand what is being said at all, then a speaker is very difficult to understand (1 below). |
| How **accented** is the speaker's speech? | This refers to how much a speaker's speech is influenced by his/her native language and/or is coloured by other non-native features. If you think the speaker has minimal or no accent, then the speaker is not accented (5 below). If you think the speaker is highly influenced by his/her native language, then the speaker is heavily accented (1 below). |

Global judgment of the speaker in terms of:

1- How difficult do you think it is to **understand** the speaker?

**Difficult to understand** **(1)** **(2)** **(3)** **(4)** **(5)** **Easy to understand**

2- How **accented** is the speaker's speech?

**Heavily accented** **(1)** **(2)** **(3)** **(4)** **(5)** **Not accented**

3- How much can you really understand the participants? For this question, write down below what you will hear:

## Appendix I: Consent Form

INFORMATION AND CONSENT FORM

Study Title: Intelligent personal assistants in L2 education: Can they help learners improve their English pronunciation?

Researcher: Souheila Moussalli

Contact Information: e-mail: souheilamoussally@hotmail.com

Faculty Supervisor: Dr. Walcir Cardoso

Contact Information: (514) 848-2424 ext. 2451 / e mail: walcir.cardoso@concordia.ca

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 students' interactions with Echo, a wireless smart speaker with built-in voice recognition that has the potential to be used for English language practice, particularly for listening and speaking skills.

B. PROCEDURES

If you participate, you will be asked to fill out a questionnaire and a survey, take some tests, and do some practice exercises that will include a series of short oral interactions with Echo (a smart speaker that you can talk to). These activities will be video- and audio-recorded. At the end of the study, you will be asked to participate in an interview which will also be video- and audio-recorded.

Your first participation will be for a period of 45 min. If you are eligible to continue with the study, your total participation in this study will take around 6 hours over a period of 4 weeks.

C. RISKS AND BENEFITS

There are no risks in participating in this research. Potential benefits include: improving overall pronunciation and vocabulary knowledge, and exposure to a new technology for language learning that promotes learning in a stress-free environment.

D. CONFIDENTIALITY

We will examine your interactions with Echo via a survey, four tests, and an oral interview as part of this research. In this study,

We will not allow anyone to access the information, except people directly involved in conducting the research. We will only use the information for the purposes of the research described in this form.

The information gathered will be identifiable. That means it will have your name directly on it.

We will protect the information by storing it on a password-protected laptop computer in a secure place in the researcher's office at Concordia University (FG 6.441). The researchers will know your name, but your name will not be revealed to anyone else. No teacher, student, or school names will be used to identify any electronic files. Codes and pseudonyms will be assigned.

We intend to publish the results of the research. However, it will not be possible to identify you in the published results.

We will destroy the information five years after the end of the study.

E. CONDITIONS OF PARTICIPATION

You do not have to participate in this research. It is purely your decision. If you 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.

There are no negative consequences for not participating, stopping in the middle, or asking us not to use your information

If you decide that you don't want us to use your information, you must tell the researcher before analysis of the study is conducted, that is before March 1 2020.

F. 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 J: Understanding the simple past tense

1. I know how to pronounce the **past -ed in English**.

   Strongly disagree 1 2 3 4 5 6 7 8 9 Strongly agree

2. In English, the **past -ed is pronounced the same in verbs such as walked, lived and invited.**

   Strongly disagree 1 2 3 4 5 6 7 8 9 Strongly agree

3. In English, **the past -ed** in **JUMPED** and **DREAMED** sound the same.

   Strongly disagree 1 2 3 4 5 6 7 8 9 Strongly agree

4. In English, **the past -ed** in **KISSED** and **JUMPED** sound the same.

   Strongly disagree 1 2 3 4 5 6 7 8 9 Strongly agree

5. In English, **the past -ed** in **PRINTED** and **DREAMED** sound the same.

   Strongly disagree 1 2 3 4 5 6 7 8 9 Strongly agree

## Appendix K: pre-test and post- test perception words

### Listening for Sounds

In this last task, you will listen to 15 words. Listen carefully and circle the sound you hear at the end of these words:

Practice:

| **word 1:** | | | **word 2:** | | |
|---|---|---|---|---|---|
| *t* | *d* | *ed* | *t* | *d* | *ed* |

Now, let's start

| | | | | | |
|---|---|---|---|---|---|
| **word 1:** | | | **word 9:** | | |
| *t* | *d* | *ed* | *t* | *d* | *ed* |
| **word 2:** | | | **word 10:** | | |
| *t* | *d* | *ed* | *t* | *d* | *ed* |
| **word 3:** | | | **word 11:** | | |
| *t* | *d* | *ed* | *t* | *d* | *ed* |
| **word 4:** | | | **word 12:** | | |
| *t* | *d* | *ed* | *t* | *d* | *ed* |
| **word 5:** | | | **word 13:** | | |
| *t* | *d* | *ed* | *t* | *d* | *ed* |
| **word 6:** | | | **word 14:** | | |
| *t* | *d* | *ed* | *t* | *d* | *ed* |
| **word 7:** | | | **word 15:** | | |
| *t* | *d* | *ed* | *t* | *d* | *ed* |
| **word 8:** | | | | | |
| *t* | *d* | *ed* | | | |

## Appendix L: pre-test and post- test perception sentences

### Listening for Sounds

In this last task, you will listen to 22 sentences. When listening to these sentences, listen carefully to the **VERB** and circle either **PAST** or **NOT PAST** based on what you heard. Then, circle which of **the three sounds** you heard at the end of these verbs.

**Let's practice this task with two sentences:**

(1) **CIRCLE** whether you heard the **PAST** or **NOT PAST**

(2) **CIRCLE** the sound you heard at the **end of the verb: t, d, or ed**

**Practice Sentence 1:**

PAST NOT PAST

*t* *d* *ed* *Ø*

**Practice Sentence 2:**

PAST NOT PAST

*t* *d* *ed* *Ø*

**Let's start:**

(1) **CIRCLE** whether you heard the **PAST tense or not**

(2) **CIRCLE** the sound you heard at the **end of the verb**: **t, d, or ed**

**Sentence 1:**

PAST NOT PAST

*t* *d* *ed* *Ø*

**Sentence 2:**

PAST NOT PAST

*t* *d* *ed* *Ø*

**Sentence 3:**

PAST NOT PAST

*t* *d* *ed* *Ø*

**Sentence 4:**

PAST NOT PAST

*t* *d* *ed* *Ø*

**Sentence 5:**

PAST NOT PAST

*t* *d* *ed* *Ø*

**Sentence 6:**

PAST NOT PAST

*t* *d* *ed* *Ø*

**Sentence 7:**

PAST NOT PAST

*t* *d* *ed* *Ø*

**Sentence 8:**

PAST NOT PAST

*t* *d* *ed* *Ø*

**Sentence 9:**

**Sentence 12:**

PAST NOT PAST

*t* *d* *ed* *Ø*

**Sentence 13:**

PAST NOT PAST

*t* *d* *ed* *Ø*

**Sentence 14:**

PAST NOT PAST

*t* *d* *ed* *Ø*

**Sentence 15:**

PAST NOT PAST

*t* *d* *ed* *Ø*

**Sentence 16:**

PAST NOT PAST

*t* *d* *ed* *Ø*

**Sentence 17:**

PAST NOT PAST

*t* *d* *ed* *Ø*

**Sentence 18:**

PAST NOT PAST

*t* *d* *ed* *Ø*

**Sentence 19:**

PAST NOT PAST

*t* *d* *ed* *Ø*

**Sentence 20:**

| | PAST | | NOT PAST | |
|---|---|---|---|---|
| | *t* | *d* | *ed* | *Ø* |

**Sentence 10:**

| | PAST | | NOT PAST | |
|---|---|---|---|---|
| | *t* | *d* | *ed* | *Ø* |

**Sentence 11:**

| | PAST | | NOT PAST | |
|---|---|---|---|---|
| | *t* | *d* | *ed* | *Ø* |

| | PAST | | NOT PAST | |
|---|---|---|---|---|
| | *t* | *d* | *ed* | *Ø* |

**Sentence 21:**

| | PAST | | NOT PAST | |
|---|---|---|---|---|
| | *t* | *d* | *ed* | *Ø* |

**Sentence 22:**

| | PAST | | NOT PAST | |
|---|---|---|---|---|
| | *t* | *d* | *ed* | *Ø* |

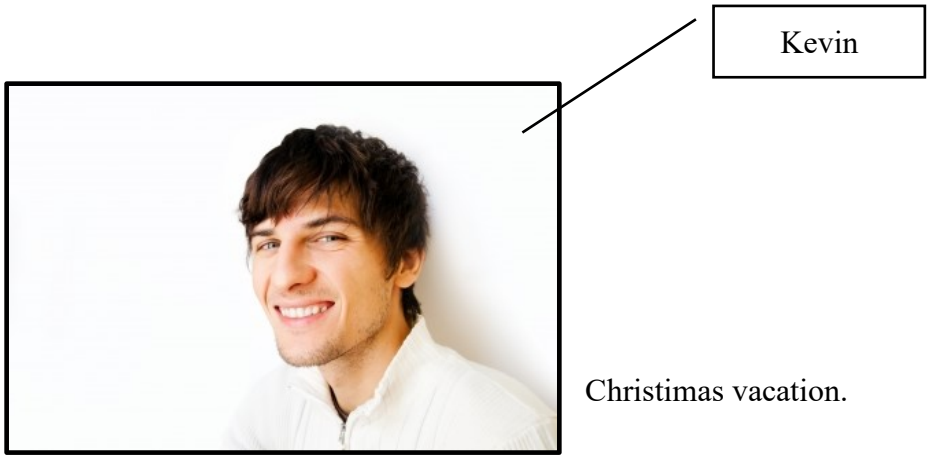
## Appendix M: pre-test production words

Read each word **followed by a long pause:**

| | | | |
|---|---|---|---|
| 1 | expected | 13 | ended |
| 2 | rushed | 14 | laughed |
| 3 | smiling | 15 | dreamed |
| 4 | passed | 16 | exploded |
| 5 | tested | 17 | followed |
| 6 | agreed | 18 | arrested |
| 7 | house | 19 | ignored |
| 8 | suspended | 20 | smelled |
| 9 | picked | 21 | excellent |
| 10 | stepped | 22 | asked |
| 11 | slowly | 23 | pencil |
| 12 | going | 24 | grabbed |

**Appendix N: pre-test controlled mock interview**

**Listen to the questions and answer following the example below:**




Kevin is planning his Christimas vacation.
**What will he do**?

**Model:**

**Q: Will he travel to France?**

| | **YES** | **NO** |
|---|---|---|
| Will he travel to **France**? | ✓ | |

YOU SAY: **Yes**, he will travel to France

| | **YES** | **NO** |
|---|---|---|
| Will he travel to **France**? | | ✓ (England) |

YOU SAY: **No**, he will travel to England

**Now, it's your turn:**

| | | **YES** | **NO** |
|---|---|---|---|
| | go to Paris? | ✓ | |
| | send postcards to his friends? | | ✓ (to his parents) |
| | speak English? | | ✓ (French) |
| Will he | travel by himself? | | ✓ (with friends) |
| | film his vacation? | ✓ | |
| | bring his girlfriend? | | ✓ (his parents) |
| | have a good time? | ✓ | |

What did Kevin do during his last summer vacation?

**Q: Did he go to France?**

| | YES | NO |
|---|---|---|
| Did he go to France? | ✓ | |

A: **Yes**, he <u>went to France</u>

| | YES | NO |
|---|---|---|
| Did he go to France? | | ✓ (England) |

A: **No**, he <u>went to England</u>

**Now, it's your turn:**

| | | YES | NO |
|---|---|---|---|
| 1 | Did he walk to the airport? | ✓ | |
| 2 | Did he travel by himself? | | ✓ (with friends) |
| 3 | Did he hate the weather ? | ✓ | |
| 4 | Did he visit his girlfriend? | | ✓ (parents) |
| 5 | Did he kiss his parents? | ✓ | |
| 6 | Did he hug his girlfriend? | | ✓ (friends) |
| 7 | Did he avoid his friends? | | ✓(ex-girlfriend) |
| 8 | Did he extend his vacation? | ✓ | |
| 9 | Did he work during his vacation? | ✓ | |
| 10 | Did he save money during his vacation? | ✓ | |
| 11 | Did he help his friends? | | ✓ (parents) |
| 12 | Did he enjoy his vacation? | ✓ | |
| 13 | Did he clean his bedroom? | ✓ | |
| 14 | Did he type e-mails every day? | | ✓ (documents) |
| 15 | Did he wash his clothes once a week? | ✓ | |
| 16 | Did he need guides to travel to France? | ✓ | |
| 17 | Did he collect souvenir? | ✓ | |

18 Did he film special moments with his parents? ✓

What else do you think Kevin did during his last vacation?

**Appendix O: Survey Learner's perception of Alexa App as a pedagogical tool**

Participant # _____

Please rate the following statements from 1 (strongly disagree) to 9 (strongly agree).

Part 1: About **LEARNING** with Alexa App (learnability)

1. Using Alexa App helped me **improve my pronunciation of past tense -ed** in English.

   Strongly disagree 1 2 3 4 5 6 7 8 9 Strongly agree

2. Using Alexa App helped me **improve my pronunciation** in English.

   Strongly disagree 1 2 3 4 5 6 7 8 9 Strongly agree

3. Using Alexa App helped me **improve** my **listening** in English.

   Strongly disagree 1 2 3 4 5 6 7 8 9 Strongly agree

4. Using Alexa App helped me **improve** my **speaking** in English.

   Strongly disagree 1 2 3 4 5 6 7 8 9 Strongly agree

5. Using Alexa App helped me **improve** my **spelling** in English.

   Strongly disagree 1 2 3 4 5 6 7 8 9 Strongly agree

6. Overall, using Alexa App helped me **improve** my English.

   Strongly disagree 1 2 3 4 5 6 7 8 9 Strongly agree

7. Overall, using Alexa App made a **positive difference** in my **English learning experience**.

   Strongly disagree 1 2 3 4 5 6 7 8 9 Strongly agree

Part 2: About **USING** Alexa App (usability)

1. I find it is **easy to use** Alexa App.

   Strongly disagree 1 2 3 4 5 6 7 8 9 Strongly agree

2. I am **comfortable** using Alexa App in English learning activities.

Strongly disagree 1 2 3 4 5 6 7 8 9 Strongly agree

3. I can **easily do** I want when I use Alexa App to learn English.

   Strongly disagree 1 2 3 4 5 6 7 8 9 Strongly agree

4. I **know how** to use Alexa App to help me **learn about pronunciation** in English.

   Strongly disagree 1 2 3 4 5 6 7 8 9 Strongly agree

5. I **know how** to use Alexa App to help me **spell** words or sentences in English.

   Strongly disagree 1 2 3 4 5 6 7 8 9 Strongly agree

6. I **know how** to use Alexa App to help me **listen** to words or sentences in English.

   Strongly disagree 1 2 3 4 5 6 7 8 9 Strongly agree

7. I **know how** to use Alexa App to help me **speak** words or sentences in English.

   Strongly disagree 1 2 3 4 5 6 7 8 9 Strongly agree

Part 3: About **MOTIVATION** to use Alexa App (motivation)

1. Using Alexa App on my own is **enjoyable**.

   Strongly disagree 1 2 3 4 5 6 7 8 9 Strongly agree

2. Using Alexa App **motivates** me to **learn about how words are pronounced** in English.

   Strongly disagree 1 2 3 4 5 6 7 8 9 Strongly agree

3. Using Alexa App **motivates** me to **listen** in English.

   Strongly disagree 1 2 3 4 5 6 7 8 9 Strongly agree

4. Using Alexa App **motivates** me to **learn about how words are spelled** in English.

   Strongly disagree 1 2 3 4 5 6 7 8 9 Strongly agree

5. Using Alexa App **motivates** me to **speak** in English.

   Strongly disagree 1 2 3 4 5 6 7 8 9 Strongly agree

6. Using Alexa App **motivates** me to **study on my own**.

   Strongly disagree 1 2 3 4 5 6 7 8 9 Strongly agree

7. Using Alexa App **motivates** me to **study** English.

   Strongly disagree 1 2 3 4 5 6 7 8 9 Strongly agree

Part 4: About my **WILLINGNESS TO USE** Alexa App (Willingness to use)

1. I would like to continue to use Alexa App as a **tool** for learning English.

   Strongly disagree 1 2 3 4 5 6 7 8 9 Strongly agree

2. I would like to continue to use Alexa App to **practice speaking** in English.

   Strongly disagree 1 2 3 4 5 6 7 8 9 Strongly agree

3. I would like to continue to use Alexa App to **practice listening** in English.

   Strongly disagree 1 2 3 4 5 6 7 8 9 Strongly agree

4. I would like to continue to use Alexa App to **practice spelling** in English.

   Strongly disagree 1 2 3 4 5 6 7 8 9 Strongly agree

5. I would like to continue to use Alexa App to **practice pronunciation** in English.

   Strongly disagree 1 2 3 4 5 6 7 8 9 Strongly agree

6. I would like to continue to use Alexa App to learn **English on my own**.

   Strongly disagree 1 2 3 4 5 6 7 8 9 Strongly agree

7. I would like to continue to use Alexa App to **learn English**.

   Strongly disagree 1 2 3 4 5 6 7 8 9 Strongly agree

## Appendix P: practice the –ed forms

Practice 1:

| Alexa, repeat after me: Simon says | |
|---|---|
| 1 | Rubbed |
| 2 | Asked |
| 3 | Avoided |
| 4 | Developed |
| 5 | tested |

| Alexa, how do you pronounce | |
|---|---|
| 1 | learned |
| 2 | blessed |
| 3 | decided |
| 4 | dropped |
| 5 | protected |

Practice 2:

| Alexa, repeat after me: Simon says | |
|---|---|
| 1 | Analyzed |
| 2 | described |
| 3 | obeyed |
| 4 | impressed |
| 5 | tested |

| Alexa, how do you pronounce | |
|---|---|
| 1 | apologized |
| 2 | grabbed |
| 3 | followed |
| 4 | kissed |
| 5 | accepted |

Practice 3:

| Alexa, repeat after me: Simon says | |
|---|---|
| 1 | passed |
| 2 | checked |
| 3 | recognized |
| 4 | longed |
| 5 | divided |

| Alexa, how do you pronounce | |
|---|---|
| 1 | Escaped |
| 2 | Scrubbed |
| 3 | Sneezed |
| 4 | Belonged |
| 5 | ended |

Practice 4:

| Alexa, repeat after me: Simon says | |
|---|---|
| 1 | dragged |
| 2 | smelled |
| 3 | dreamed |
| 4 | dressed |
| 5 | laughed |

| Alexa, how do you pronounce | |
|---|---|
| 1 | walked |
| 2 | Traveled |
| 3 | saved |
| 4 | crossed |
| 5 | included |

## Appendix Q: Interview questions

1. Tell me about your experience using Echo.
2. What are its strengths? What are its strengths as a tool to learn and practice English?
   a. Did you find ECHO helpful/ beneficial? In what ways?
3. What are its weaknesses? What are its weaknesses as a tool to learn and practice English?

   a. Did you have any problems while working with ECHO? What kind of problems? Were they resolved easily? How?
   b. Did you have to repeat many times? How many times? Can you give me a number?
   c. Did you find repeating helpful? If no, why?

4. Overall, did you enjoy (or not enjoy) working with ECHO? Why?
5. Would you ever consider using ECHO again for learning a language? Why?
6. Was Echo able to let you know when you mispronounced a word? For example, by asking you to repeat the question, by giving you a different answer, etc.? Was Echo's way of dealing with your mispronunciation useful?
   a. Did you find repeating helpful? If no, why?
7. Did you understand ECHO/Alexa? Do you like the voice? Is it fast?
8. Did Echo/ Alexa understand you?
9. What strategies or steps did you take to do the activities?
10. How many times did you listen to the texts and listening?
11. How much time did you spend on the activities?
12. Do you think Alexa helped you pronounce the past tense better?

13. Would you use ECHO again for other purposes?

- Opportunities to **listen**, to practice listening
- Opportunities to **speak**, to practice speaking
- Opportunities to **learn about pronunciation**
- Opportunities to **repeat**, to practice what you're learning
- It gives you **feedback** (for example, by responding to you when it understands you)
- To have **someone** to talk to, who speaks and understands English
- For **motivation**
- For **fun**

## Appendix R: Practice with Alexa -ed sounds

ECHO questions

**Language**:

1. Alexa, spell _____
2. Alexa, ask pronunciations how to pronounce L.A.U.G.H.E.D
3. Alexa, ask pronunciations to pronounce A.S.K.E.D
4. Alexa, open Translated. Ask translated: 'love in French" (translates into 50 languages*)
5. Alexa, repeat after me: _____
6. Alexa, what is the definition of _____?
7. Alexa, what is the antonym of _____?
8. Alexa, tell me a story.
9. Alexa, open interactive stories.
10. Alexa, open Dr. Speech.
11. Alexa, read me a poem.
12. Alexa, say the alphabets
13. Alexa, ask TED TALKS to find talks about _____
14. Alexa, open grammar tool.

**General information: math- science- geography- history- Wikipedia-chemistry**

1. Alexa, how tall is Mount Everest?
2. Alexa, what is the capital of_____?
3. Alexa, how far is the moon?
4. Alexa, what is 2 +2?
5. Alexa, what is the square root of 64?
6. Alexa, How many cups in a litre?
7. Alexa, Wikipedia: " Niagara Falls"
8. Alexa, what is the definition of _____?
9. Alexa, ask Today in History about a famous birthday today.
10. Alexa, ask Today in History to tell me a random fact about today.

**Games: do you know any of the games?**

1. Alexa, tell me a joke.
2. Alexa, tell me a riddle.
3. Alexa, play rock paper scissors
4. Alexa, knock knock (enable skill if doesn't work*)
5. Alexa, Simon says: **KISSED**
6. Alexa, play Jeopardy.