

EXPERIENCES OF SPEAKING  
WITH CONVERSATIONAL AI IN  
LANGUAGE EDUCATION



EXPERIENCES OF SPEAKING  
WITH CONVERSATIONAL AI IN  
LANGUAGE EDUCATION

ELIN ERICSSON

Doctoral Dissertation

Department of Applied Information Technology  
University of Gothenburg  
SE-412 96 Gothenburg  
Sweden

© Elin Ericsson

ISBN 978-91-8069-351-6 (PRINT)

ISBN 978-91-8069-352-3 (PDF)

Doctoral Thesis in Applied Information Technology towards Educational Science,  
at the Department of Applied IT, University of Gothenburg.

This thesis is available in full text online

<http://hdl.handle.net/2077/77123>

This doctoral thesis has been prepared within the framework of the graduate school  
at the Centre for Educational and Teacher Research, University of Gothenburg.

Centre for Educational Science and Teacher Research, CUL

Graduate school in educational science

Doctoral thesis 104

In 2004, the University of Gothenburg established the Centre for Educational Sciences and Teacher Research (CUL). CUL aims to promote and support research and third-cycle studies linked to the teaching profession and the teacher training programme. The graduate school is an interfaculty initiative carried out jointly by the faculties involved in the teacher training programme at the University of Gothenburg and in cooperation with municipalities, school governing bodies and university colleges.

[www.cul.gu.se](http://www.cul.gu.se)

Cover photo showing the work of art *The Conversation* by the sculptor Etienne.  
Photo by Elin Ericsson, Department of Applied Information Technology  
Division of Learning, Communication, and IT.

All papers included in this thesis have been printed with permission.

Printed in Borås, Sweden 2023

Printed by Stema Specialtryck AB



## ABSTRACT

This thesis explores the application of digital tools in Swedish language education, specifically in relation to speaking skills. It focuses on dialogue-based, computer-assisted language learning, which enables students to practise and develop their speaking skills in a target language. The aim of the thesis is to gain a comprehensive understanding of how students experience conversational artificial intelligence during their language education while practising speaking skills in face-to-face interaction.

The thesis consists of one study conducted with language teachers, and three studies conducted in real classroom settings with lower-secondary students using two different spoken dialogue systems. A framework was designed to analyse student–conversational-agent interaction and the students' educational experiences in the system. The data was produced through system metrics, questionnaires, digital logbooks, and interviews. Both descriptive and inferential statistics were employed to analyse the quantitative data, while qualitative data was analysed using reflexive thematic analysis.

There were three major areas of findings related to the conducted studies. Firstly, teachers have positive views of using digital tools, but speaking skills are the least assisted by these tools, despite being considered the most challenging skill to teach and learn. Secondly, students reported positive educational experiences when practising speaking in a spoken dialogue system, both in the short and long term. They were cognitively and emotionally engaged, felt safe, and gained confidence. However, some students became disengaged when the scenarios did not relate to their daily lives, lacked challenge, or resulted in communication breakdowns. Thirdly, students had a range of experiences with the embodied conversational agents, from seeing them as deadpan machines to relating socially with them in positive ways. The level of experienced social interaction was linked to the students' educational experiences in the system.

The contributions of the thesis include the framework for analysing student–conversational-agent interaction and the adaptation of a digital logbook for this educational context. The practical implications for language education are various. Firstly, spoken dialogue systems offer supplementary opportunities for lower-secondary students to practise speaking

in a target language and align with key principles of effective language learning. Secondly, understanding the complexity of various aspects of students' experiences in the system can help teachers to address challenges and diverse student reactions, transforming their role into that of a facilitator and guide for teaching and learning speaking skills using conversational AI.

**Keywords:**

Conversational artificial intelligence; dialogue-based computer-assisted language learning; educational experience; embodied conversational agent; student–conversational-agent interaction framework; lower-secondary students; speaking skills; spoken dialogue system; teachers.

## PREFACE

The cover photo of this thesis represents two, non-identifiable interacting characters. It was taken in the heart of Havana in Cuba in 2014, during my eye-opening travels there, which I have documented as “*Diez días en Cuba Bailando*” [Ten days dancing in Cuba]. This photo has acted as my guiding star, always present in my public presentations during this doctoral journey, as a symbol of spoken dialogue and the importance of interaction. Imagine replacing one of these characters with a Swedish teenager and the other with an embodied conversational agent, and you will have the setting for this thesis. If you love languages, set them free!





## ACKNOWLEDGEMENTS

This thesis was designed and produced within two parallel research school environments, where teachers and colleagues provided unique opportunities for sharing and developing knowledge and skills together. These environments have contributed to “*Bildung*” [enlightenment and humanistic education], referring to the process of my personal and cultural development through this education. Firstly, I was admitted to the interdisciplinary graduate school in educational science at the Centre for Educational Science and Teacher Research (CUL) in Gothenburg in Sweden, on the Applied Information Technology programme. Secondly, I was admitted to the interdisciplinary national Graduate School - Digital Technologies in Education (GRADE).

The idea for this thesis sprang out of my genuine interest in languages and the joy of being able to communicate with people in their own language. I learnt English, French, and some German at an early age and Spanish as an adult. Working as a language teacher for more than ten years and as a development manager of digitalisation within education, I have developed a passion for teaching and learning. My research interest in speaking skills emerged from private and professional experiences of speaking as challenging to teach and practise in a classroom context, although it is still the most frequently used skill in my learnt languages applied in everyday-life situations. With today’s opportunities for using digital tools employing conversational AI, which may be used in a language classroom context as a supplementary method for practising speaking one’s target language, I wanted to explore their potential, still with the assumption that human-to-human conversation remains the gold standard and end-goal of language learning. Even though this research was conducted and presented in this thesis by one author and defended by the same person, it is the product of an outstanding group of different people who have supported me and contributed in various ways along this learning journey. Without mentioning all of you here by name, you know who you are, and I want to give you my tremendous thanks for giving me your sustained interest and engagement.

Firstly, I would like to express my deepest gratitude to my main supervisor, Johan Lundin, who believed in me from day one, guiding me for-

ward on this creative research journey together with my second supervisor, Sylvana Sofkova Hashemi, who shared my interest in language learning and supported my SLA key principles against strong winds, and, finally, Anders D. Olofsson from the north, who brought new energy to the team. I can hear my new version of Status Quo singing in chorus: “We are in the *process* [army] now, oh oh oh, we’re in the *process now*”, convinced that this team will make it to the end! A special thank you also goes to Stefan Johansson, for generously sharing your expertise in supervising the analysis of the longitudinal data that constituted the grand finale of my thesis. I am also indebted to Oskar Lindwall, for being an objective examiner.

Tore Nilsson, my warmest thanks for enriching my research with invaluable discussions about second-language teaching and learning, especially speaking skills, and your generosity in sharing your expertise within the field.

This research would never have seen the light of day without all the teachers and students who participated in the studies with patience and curiosity about new emerging technologies. In particular, I would like to thank *The* teacher, who enabled the student studies (not mentioned by name due to ethical guidelines) in a lower-secondary school (grades 7 and 8). My sincere thanks are also due to The Spanish Teacher Association (*Asociación de Profesores de Español del Oeste de Suecia*, APEOS) and The Language Teachers’ Association (*Språklärarnas Riksförbund*, SPRI), which provided the springboard for this thesis.

Thank you to *all* my colleagues for welcoming me into the community at the Department of Applied Information Technology (AIT), especially to some former and present doctoral colleagues at the Division of Learning Communication and IT, who have generously shared particularly memorable moments with me: Anita, Charlotte A, Christina, Karin, and Marie U M. Initially, Lisa and Martin enthusiastically shared the long-awaited acceptance notice with me and participated in our unforgettable writing retreat at an Italian monastery, setting the lovely stage for this period of my life as a doctoral student. Charlotte A has been a continuous source of encouragement and optimism throughout this research journey (read rollercoaster). We have a lot in common, such as reflection and engagement in our conversations, expressed preferences for bathing, and believing in the correlation of creativity and water, as well as learning in interaction.

Thank you, Charlott S and previously Johan, for your support in your roles as Head of Division. On a similar note, Helena, the present Head of Department and the previous one, Jonas, thank you for your support. I am also immensely grateful to Mattias for assisting me with the material for the studies and in making this research printable and publishable. A big thank you to Pär for the warm assistance given from day one to the very end of this thesis. Thank you, Jenelyn and Karin, for your careful administrative assistance. My special thanks go to Marie E for your warmth and encouragement. Thank you Ylva for the wonderful metaphor of the life of a PhD student, developing from a pupa into a flying butterfly.

The CUL network has given me the opportunity to meet numerous colleagues who share the same passion for educational science with recurring annual meetings both within and outside the institutional context. Thank you Henrik for introducing us into the philosophy of knowledge. A special thanks goes to Minna for sharing beneficial and challenging moments together focusing on statistics.

The GRADE research school network has given me the opportunity to meet wonderful colleagues who share the same interests in digitalisation. My gold medal for the boule tournament reminds me of valuable shared moments and discussions, both formal and informal. Thank you, Malin, Stefan, and Stefan for sharing your expertise and inspiring my design of the framework of student–conversational-agent interaction (FoSCAI) in relation to language learning and development. Karoline has provided me with nice company and doctoral discussions at a distance during the Pandemic, sharing Zoom coffee breaks.

I must express my sincere appreciation to my seminar opponents for their careful reading and feedback, which led me forward in the research. Roger Säljö, thank you for initially giving me the right wind in my sails, which led me in the right direction. Thank you, Jimmy Jaldemark, for highlighting concepts mid-way through the process, such as interaction. Ola Knutsson, thank you for the final reading of this text and confirming “The grand finale” of the narrative.

My heartfelt thanks go to Jenny Mattsson and Richard La Bontee at The Unit for Academic Language (ASK) for coaching me in the development of academic language and writing, including in this “text jacket” (kappa).

Historically, right through from my language teacher education period at Gothenburg University, there has been a multitude of teachers who have encouraged and inspired me to become a doctoral student, such as my former supervisors Mats Mobärg, Marianne Molander Beyer, and Iah Hansén.

Tina, I am grateful for the support extended towards my aspirations and the provision of leave of absence from my employment position.

I want to express my gratitude to Martin Lackéus and Carin Sävetun, who have opened the door to the endless possibilities of using their digital logbook LoopMe for systematic data production over time.

Robin Teigland, thank you for the inspiration: “If we are going to move forward, we have to get closer to technology” and “walk the line between theory and practice”. In this thesis, the focus is not on enabling marine communities to develop new ocean possibilities for an overfished surfing coastal community in Portugal that is to be rejuvenated, as in her speech, but rather on conversational AI providing potentially new interactional possibilities for supporting the learning and development of L2 speaking for students.

Internationally, I owe a great debt of gratitude to my unofficial supervisor, Regina Kaplan-Rakowski, for introducing me to research on virtual reality while lecturing in Sweden and to the international Computer-Assisted Language Instruction Consortium (CALICO). I have actively used my speaking skills in French, Spanish, and English when participating in different international conferences, such as CALICO in the USA, and ICERI, EDULEARN, INTED, and ICERI in Spain, where I met and interacted with many wonderful people. My gratitude is extended to Serge Bibauw, who inaugurated me into the subfield of dialogue-based CALL through our computer-mediated dialogues on Zoom, me in the summer-like Swedish countryside and him in Ecuador. I am forever grateful to Rose Luckin for enabling the contact with Alelo Inc. and Lewis Johnson, who welcomed my research and gave me and the recruited students access to their spoken dialogue system, Enskill. I wish to express my sincere appreciation to Scott Cardwell for letting me use his spoken dialogue system ImmerseMe with students while conducting this research.

Most of all, I am extremely grateful to my family, who have supported me with endless love and patience throughout these intense years while this


research seeped into our lives and sometimes caused me to drift away in long and late working sessions. Mother and Father, nothing of this research would have been possible without you. Thank you for always being there for me with your generosity, love, support, and encouragement. Profound gratitude goes to my Mother for endless mental support and reading of my texts, followed by invaluable discussions and engagement in my research. My warmest thanks to Amanda for sisterly and professional supervising, and to my brother-in-law Paul for your reading of manuscripts and discussing linguistic nuances. Thanks also to my brother Jonas for providing me with successful moments enriched with delicious tea. I am particularly grateful to my Jaime for sharing many wonderful moments, boosting me with energy, conversations in Spanish, love, and salsa dance.

Cuban music has kept me moving forward through this doctoral journey, although some days have been more of a challenge than others. A parallel can be drawn with the frequently repeated famous rhythmic danceable phrase, the metaphorical signum of Havana de Primera, sung with so much energy by Alexander Abreu in “*Vaya, camina por arriba el mambo*” [Oh, walk on top of the mambo], expressing *strength and power* in movements forwards and backwards. In this context, I interpret it as keeping the rhythm and energy flowing, and sometimes helping me to walk on top of things, turn around, go back some steps and restart the advance again later. Achieving this long-term goal has required a combination of persistence, resilience, and passion to make it through the difficult phases, but this perseverance has helped me to succeed with this thesis.

Friends have been a continuing source of support and encouragement along the way. My warmest thanks go to Annika F, who has always been there for me, listening and sharing everything. Also to Andreas for nice walks and talks by the sea, which gave me fresh air and new insights in order to meet deadlines. Percy and Cecilia have been there for me and the whole family, spreading joy and support throughout these years. Maria and Niklas, thank you for sharing your passion for learning and giving me a final rewarding target worth working extra hard for (yes, the glass is now heavy and full, and we are beautiful together again in town). Abbe and Malin, thanks for your patience with my absence. Thanks to Sabrine and Carine for keeping up my ability to communicate in French while following my research at a distance from Mediterranean latitudes and the

French Alps, respectively. Annika L, our summer meetings in real life and interactions across the ocean have been vital for me in this process.

Last but not least, I dedicate this thesis to my daughters, Elvira and Felicia, for all your patience and love throughout this doctoral journey. Inspired by the French singer Francis Cabrel, singing about eternal love, I address this line to you: “*Je vous aimais, je vous aime et je vous aimerai*” [I loved you, I love you, and I will love you]. Children represent the next generation, who will hopefully benefit from conversational AI when learning and developing their language skills in a target language. Speaking is at the heart of language teaching and learning. Being able to speak different languages has been a weightless key to many interesting meetings with new people and much joy for me, in both working and private life. I would like to share that extra dimension of life and facilitate for language students (and life-long learners) through this research. As Steven Thorne expressed it in an unforgettable face-to-face conversation about CALL and my research in Seattle in May 2022: “The best PhD thesis is the done thesis”, and *voilà, bitte schön, toma*, here IT is!

Elin Ericsson, Gothenburg, June 2023 

## ADDITIONAL PAPERS RELATED TO THE THESIS

Ericsson, E. (2022). Digital logbook for producing real-time longitudinal data. In *INTED Conference Proceedings of the 16th International Technology, Education and Development Conference* (pp. 7747–7756). IATED.

[https://www.researchgate.net/publication/359475495\\_DIGITAL\\_LOGBOOK\\_FOR\\_PRODUCING\\_REAL-TIME\\_LONGITUDINAL\\_DATA](https://www.researchgate.net/publication/359475495_DIGITAL_LOGBOOK_FOR_PRODUCING_REAL-TIME_LONGITUDINAL_DATA) [Full paper].

<https://doi.org/10.21125/inted.2022.1961> [Abstract]

Ericsson, E. (2023). Model of educational experience in spoken dialogue systems for language learning. In *EDULEARN23 Conference Proceedings of the 15th International Conference on Education and New Learning Technologies*. IATED. <https://doi.org/10.21125/edulearn.2023.2196>

Ericsson, E., & Jansson, M. (2021). Framework for studying students' experiences of practising foreign language speaking skills with conversational agents in spoken dialogue systems [abstract]. In *ICERI Conference Proceedings of the 14th International Conference of Education, Research and Innovation* (p. 1551). IATED. <https://doi.org/10.21125/iceri.2021.0418>

Ericsson, E., & Johansson, S. (2023). Sustained speaking experiences with virtual humans over time. In *INTED Conference Proceedings of the 17th International Technology, Education and Development Conference* (pp. 6011-6022). IATED.

[https://www.researchgate.net/publication/369579604\\_SUSTAINED\\_SPEAKING\\_EXPERIENCES\\_WITH\\_VIRTUAL\\_HUMANS\\_OVER\\_TIME](https://www.researchgate.net/publication/369579604_SUSTAINED_SPEAKING_EXPERIENCES_WITH_VIRTUAL_HUMANS_OVER_TIME) [Full paper].

<https://doi.org/10.21125/inted.2023.1589> [Abstract]





## DISTRIBUTION OF WORK

In Papers I–V, the guidelines provided in the Vancouver recommendations (ICMJE, 2023) were followed, acknowledging authorship and defining roles and responsibilities promoting uniformity, clarity, and transparency in writing these papers. There were no ethical conflicts of interest for any of the authors.

### **The following people contributed to the papers that are part of this thesis:**

Elin Ericsson<sup>1</sup> (Author 1)  
Johan Lundin<sup>1</sup> (Author 2)  
Sylvana Sofkova Hashemi<sup>2</sup> (Author 3)  
Stefan Johansson<sup>3</sup> (Author 4)

<sup>1</sup>Department of Applied Information Technology, University of Gothenburg, Sweden

<sup>2</sup>Department of Education, Communication, and Learning, University of Gothenburg, Sweden

<sup>3</sup>Department of Education and Special Education, University of Gothenburg, Sweden

### **The roles of the authors:**

#### **Paper I**

As the primary author (Author 1), I was responsible for the comprehensive work around the study's idea, design, data production, analysis, and empirical work in conducting Study 1. I created the questionnaire and refined it after input from teachers in Pilot 1 and supervised by Authors 2 and 3. I wrote the manuscript, to which Authors 2 and 3 contributed to content and analysis, before subsequently approving the final version and agreeing to be held accountable for the work. I presented the work at an academic conference.

## **Paper II**

As the primary author (Author 1), I was responsible for the comprehensive work around the study's idea, design, data production, analysis, and empirical work in conducting Study 2. I created the questionnaire and logbook under the supervision of Authors 2 and 3. I wrote the manuscript, to which Author 2 contributed to content and analysis, and 3 to refinement, before Authors 2 and 3 subsequently approved the final version and agreed to be held accountable for the work. I presented the work at an academic conference.

## **Paper III**

As the primary author (Author 1), I was responsible for the comprehensive work around the study's idea, design, data production, and empirical work in conducting Study 3. I created the questionnaire and logbook under the supervision of Author 3, who also participated in coding during the thematic analysis. I wrote the first draft of the manuscript, to which Author 3 contributed content and analysis, and participated in the revision processes before approving the final version and agreeing to be held accountable for the work. I presented the work at an academic conference.

## **Paper IV**

As the primary author (Author 1), I was responsible for the comprehensive work around the study's idea, design, data production, and empirical work in conducting Study 4. I created the framework FoSCAI, and developed and designed the relevant questionnaires and logbook under the supervision of Authors 2 and 3. I wrote the first draft of the manuscript, to which Author 2 contributed with content and analysis. Author 3 participated in the coding and refinement of the thematic analysis. Authors 2 and 3 approved the final version and agreed to be held accountable for the work. I presented the work at an academic conference as well as FoSCAI with a GRADE colleague at another academic conference.

## **Paper V**

As the primary author (Author 1), I was responsible for the comprehensive work around the study's idea, design, data production, and empirical work in conducting Study 4. I created and refined FoSCAI, developed and

designed the relevant questionnaires and logbook. I conducted the analysis in consultation with Author 4, who assisted in conducting statistical operations and presenting results. I wrote the first draft of the manuscript, to which Author 4 contributed with refinements, approved the final version, and agreed to be held accountable for the work. I presented the work at an academic conference.



# CONTENTS

## PART 1 – EXPERIENCES OF SPEAKING WITH CONVERSATIONAL AI IN LANGUAGE EDUCATION

INTRODUCTION .....	29
1.1 Second-Language teaching and learning – speaking	34
1.2 Computer-assisted language learning – speaking	37
1.3 Teachers’ views on and experiences of digital tools	38
1.4 Students’ experiences of conversational AI	39
1.5 Aim and research questions	41
1.6 Outline of this thesis	41
LANGUAGE EDUCATION IN SWEDEN .....	43
2.1 Studying languages	43
2.1.1 Modern languages and English	45
2.1.2 Communicative and action-oriented approach	46
2.2 Digitalisation in Swedish (language) education	51
CONVERSATIONAL AI .....	53
3.1 Spoken dialogue systems and virtual agents	53
3.1.1 Embodied conversational agents	56
3.2 Past, present, and future	57
3.2.1 Application in language education	58
DIALOGUE-BASED, COMPUTER-ASSISTED LANGUAGE LEARNING – SPEAKING SKILLS.....	61
4.1 Dialogue-based, computer-assisted language learning	61
4.2 Second-language acquisition and computer-assisted language learning	64
4.3 Interplay of cognitive and affective factors	68
4.4 Principles for effective L2 learning	69
4.4.1 Authentic input and focus on meaning in interaction	70
4.4.2 Noticing of form and feedback	71
4.4.3 Low affective filter	72

RELATED RESEARCH .....	75
5.1 L2 Teachers' views and experiences	75
5.1.1 Studies in relation to the employed SDSs	78
5.2 Students practising L2 speaking skills in an SDS	79
5.2.1 Beneficial for L2 speaking, listening, and vocabulary	79
5.2.2 Low-anxiety environment, gains in confidence and willingness to communicate	81
5.2.3 Effectiveness versus bias	81
5.2.4 Attitudes, engagement, and motivation for L2 learning	83
5.2.5 Goal-oriented systems, feedback, and scaffolding	84
5.2.6 VR immersion in SDSs	84
5.2.7 Studies in relation to the employed SDSs	85
5.3 Experiences of L2 speaking with CAs	87
5.4 Relevance to this thesis	90
 THEORETICAL UNDERPINNINGS .....	 93
6.1 Communication and social interaction	93
6.1.1 Interaction hypothesis	96
6.2 Learning by speaking the target language	97
6.3 Experience	98
6.4 Framework of student–CA interaction	99
6.4.1 Presence, experience, and interaction	99
6.4.2 Dimensionalising of experience	102
6.4.3 Educational experience	106
 RESEARCH DESIGN AND METHODS .....	 107
7.1 Research design	107
7.1.1 Experience as an analytical unit	109
7.1.2 Participants	110
7.1.3 Setting and realisation of Studies 2–4	111
7.1.4 Selection of spoken dialogue systems	114
7.2 Data production	125
7.2.1 Metrics	126
7.2.2 Teacher questionnaire	126
7.2.3 Student questionnaires	127
7.2.4 Digital logbooks – shared document and LoopMe	130
7.2.5 Semi-structured interviews	132
7.3 Methods and data in studies 1–4	133
7.3.1 Pilot study	135
7.3.2 Study 1	135

7.3.3 Study 2	136
7.3.4 Study 3	136
7.3.5 Refinement before Study 4	137
7.3.6 Study 4	138
7.4. Data analysis	139
7.4.1 Study 1	141
7.4.2 Studies 2–4	141
7.4.3 Study 4	142
7.5 Ethical concerns	143
7.6 The role of the teacher and the researcher	145
SUMMARY OF PILOT AND PAPERS.....	147
8.1 Pilot	147
8.2.Paper I	148
8.3 Paper II	150
8.4 Paper III	152
8.5 Paper IV	156
8.6 Paper V	158
CONTRIBUTIONS AND DISCUSSION .....	163
9.1 Language teachers ready to implement SDS (RQ1)	164
9.2 Fun, safe, and frustrating – practising L2 speaking (RQ2)	165
9.3 From deadpan machine to social interaction (RQ3)	169
9.4 Framework of student–CA interaction and logbook	171
9.5 Methodological reflections	173
9.5.1 Spoken dialogue systems and languages	173
9.5.2 Limitations	175
9.5.3 Self-reporting, trustworthiness and instruments	176
9.5.4 Validity and reliability	177
9.6 Practical implications	178
9.6.1 Alignment with SLA principles	178
9.6.2 The complexity of practising L2 speaking in an SDS	179
9.6.3 Changes in the educational landscape	180
9.6.4 Suggestions for improvements in SDSs	181
9.7 Future research	182
CONCLUSION .....	185

SWEDISH SUMMARY.....	187
11.1 Introduktion	187
11.2 Teori och metod	189
11.3 Dataanalys	190
11.4 Genomförda studier	190
11.5 Resultat och bidrag	192
REFERENCES .....	195

## PART 2 – THE PAPERS

### PAPER I

Teachers' attitudes and experiences of digital tools within speaking proficiency in foreign teaching and learning – Interactional competence left in the dark

### PAPER II

Fun and frustrating: Students' perspectives on practising speaking English with virtual humans

### PAPER III

(Dis)engagement and usefulness: Students' experiences of practising speaking German in a virtual 360° environment

### PAPER IV

From deadpan machine to relating socially: Middle-school students' experiences speaking English with embodied conversational agents

### PAPER V

English speaking practice with conversational AI: Lower-secondary students' educational experiences over time



## ACRONYMS

AI	Artificial Intelligence
AIED	AI in Education
ASR	Automatic Speech Recognition
CA	Conversational Agent
CALL	Computer-Assisted Language Learning
CEFR	Common European Framework of Reference for Languages
DB-CALL	Dialogue-Based Computer-Assisted Language Learning
ECA	Embodied Conversational Agent
FoSCAI	Framework of Student–CA Interaction
L2	Second Language
L3	Third Language
NLP	Natural Language Processing
SDS	Spoken Dialogue System
SLA	Second-Language Acquisition
TBLT	Task-Based Language Teaching
RQ	Research Question
VH	Virtual Human
VR	Virtual Reality
ZPD	Zone of Proximal Development



PART 1

---

**EXPERIENCES OF  
SPEAKING WITH  
CONVERSATIONAL AI IN  
LANGUAGE EDUCATION**



# CHAPTER 1

## INTRODUCTION

It has been argued that artificial intelligence (AI) represents the most significant paradigm shift ever seen in relation to the ability of machines and systems to perform tasks that typically require human intelligence (Tewari, 2022). AI transforms how we learn, work, research, live, and think, pushing humans towards lifelong learning and reinventing old habits (Tegmark, 2017; Teigland, 2018). Making use of the progression of AI is expected to change future jobs and education and it has been found that Swedish teachers are interested in exploring the opportunities and challenges of AI in education (AIED) but needed to gain more insights to use its full potential (Hrastinski et al., 2019). As Selwyn (2023a) has proposed, the “coming of age of AIED” (p. 1) may already be here, and he argues for proponents and critics to work together to avoid a polarised debate (see the nuanced reply of scholars in Åkerfeldt et al., 2021). Dawson et al. (2023) have raised the concern that while AI is revolutionising education at high speed, there is a crucial need for theoretical frameworks in order to guide research and practice. According to Utterberg Modén et al. (2023), AIED changes educational practice, and one example is emerging automated decision-making technologies for assessment (Cerratto Pargman et al., 2023). Digitalisation changes the conditions for learning, and this affects students’ engagement, which Bergdahl et al. (2020) found to be higher among high-performing upper-secondary school students than low-performing students. Since engagement is critical for learning, more insights are called for to understand the needs of different student groups. AIED not only holds out promises, such as personalised learning trajectories and teacher assistance, but also poses threats due to poor implementation and weaknesses (Humble and Mozelius, 2022), in aspects such as

missing ethical guidelines (Figueras et al., 2022). One strength of AIED concerns the subfield of natural language processing (NLP), which can be used beneficially in the area of language education (Clark, 2023; Humble & Mozelius, 2022). This thesis explores language education in Sweden related to experiences of AI use, with a specific focus on speaking.

One kind of AI that is of particular interest is conversational AI (McTear, 2020), which is generally used as an umbrella term for speech technologies that enable conversations with computer-controlled artificial interlocutors, referred to as conversational agents (CAs), for practising the target language (Bibauw et al., 2019; Craig & Schroeder, 2018). CAs range from voice- or text-based chatbots (Huang et al., 2021; Hwang & Chang, 2021) to physical agents such as robots and spoken dialogue systems (SDSs). SDSs can provide a low-anxiety virtual environment peopled with humanlike CAs (Ayedoun et al., 2015), also referred to as embodied CAs (ECAs), to promote face-to-face social interaction (Bickmore & Cassell, 2005; McTear, 2020). SDSs enable immersion by means of sounds and images (Godwin-Jones, 2021) in “culturally authentic” everyday-life scenarios (Berti, 2019, p. 327). Dialogues are here about solving a task by actively using the target language (Ellis, et al., 2020) in scenarios revolving around activities such as ordering food at a restaurant or making plans to attend a music festival (Johnson, 2019a).

Conversational AI used for communication and interaction is rapidly becoming a part of society in commercial products such as voice-enabled speakers in Siri (Bellegarda, 2014) and Google Home Assistant (n.d.; Google Duplex, 2018), and various customer-support services online. In fact, the development of modern AI and machine learning made it possible over 60 years ago for humans to interact *with* intelligent machines and systems, where the computer is the interlocutor (Turing, 1950; Weizenbaum, 1996; Winograd, 1972). Such interaction today can be visual, written, oral, or a combination of these, and synchronous (real-time) or asynchronous (deferred time). In this thesis, the terms dialogue and conversation are used interchangeably since they are already part of the applied terminology (e.g., ECA and SDS).

In institutional language education and digital teaching materials and tools in Sweden, conversational AI has so far been employed to only a limited extent, if at all. One example from practice was recently highlighted in

the Swedish news, which showed an educational situation in an AI-based virtual reality (VR) environment for students to practise “speaking with people” in the target language (Swedish Public Service Television, 2023, March 18). The ECAs were visualised through head-mounted devices. In this news report, it appeared that the experienced benefits of this educational technology were frequent practising, with “questions thrown at you non-stop”, which stimulated speaking practice. Additionally, students who normally felt too shy to speak in front of others in the class were now able to speak. In recent years, Swedish educational policy documents have strengthened the requirements for the development of teachers’ and students’ digital competence and the use of digital tools (National Agency for Education, 2018a, 2018b). However, the area of pedagogically sound inclusion of digital tools in relation to language education in a Swedish context is under-researched (Nilsson, 2017).

Despite the increased availability of conversational AI, and its recognition in language education, relatively few studies have been conducted in this area (Ji et al., 2022). Those that do exist have typically focused on students in higher education, evaluating and measuring the effects of SDSs on learning (Bibauw et al., 2022) and linguistic performance, but rarely specifically on speaking skills and interaction (Bajorek, 2017; Lin & Lan, 2015). Consequently, few studies have been found that explore lower-secondary school students’ experiences of speaking practice in an SDS in relation to the learning and development of speaking skills. In the myriad of individually experienced aspects in an SDS, both internal and external, that might bear on meaningful practice opportunities and eventually effective learning within SDSs, many remain to be explored (Morton et al., 2012; Timpe-Laughlin et al., 2022). For instance, how students relate socially to ECAs is still an under-researched aspect, which is important if SDSs are to be used as a complementary method for practising speaking. Fryer et al. (2020) suggest a more efficient use of CAs in language education after further adjustment of the agents in collaboration between researchers and developers. Additionally, few SDSs have made it beyond prototyping to become available to the public or for education (Bibauw et al., 2019).

In this thesis, conversational AI is represented through the two SDSs, Enskill (Alelo, 2022) and ImmerseMe (n.d.). They include two separate

types of ECA: (i) animated virtual humans in an animated environment and (ii) pre-filmed human beings in 360-degree videos. ECAs in both types of system include automatic speech recognition (ASR) technologies, allowing students to speak and make themselves understood. The intention is to investigate how these systems enable students to engage in meaning-focused, face-to-face dialogues in the target language in a virtual environment, without evaluating or comparing the systems per se. The students receive instant feedback from the ECA to varying degrees depending on the system's design, which has implications for the spectrum of constraints in the dialogues and pedagogical possibilities (Bibauw et al., 2019). The differences between the systems were reflected in the students' experiences of practising speaking in them. Corrective feedback, followed by individual assessment based on errors in the spoken utterances of the turn-taking, may be provided.

The term "practise" is used in line with Ellis et al. (2009) in skill-based language teaching, where learning and developing the ability to speak (implicit knowledge) need to be practised through spoken communication (speaking), as opposed to learning aspects such as grammar and vocabulary isolated in form-exercises (explicit knowledge). The participating students practised interacting with an ECA individually in the target language, as exemplified in the quoted dialogue between the English-speaking virtual human Tony and a Swedish student (Figure 1.1).



**Figure 1.1**

*Transcript excerpt of spoken English interaction between the virtual human Tony and a Swedish student practising speaking in Enskill*



-Hey, how are you?

-There's a music festival coming to town. Do you want to go?

-Which day do you want to go?

-Why do you want to go on Saturday?

-What kind of music do you like?

-Why do you like pop best?

-Cool! Me too!

-\$100 a day. Is that okay?

-Great!

-The festival is at eight o'clock.

-Do you want me to pick you up?

-Well, that's all I know about the festival.  
I'll call you before the weekend. See you!

-Hello Tony!

-I'm good, thanks for asking. How are you?

-Yeah, of course, I want to go. When is the festival?

-I think I would like to go on Saturday

-Because I like the day better.

-I like pop and hip-hop, and what about you?

-Cause it makes me happy, and I want to dance when I hear it.

-How much is the ticket?

-Yeah, that's ok!

-What time is the festival on Saturday?

-How will we get to the festival?

-Yeah, that sounds great!

-See you, goodbye!

*Note.* © Alelo Inc. Used with permission.

Tony and the student took turns, and the dialogue proceeded as presented in this system-transcribed excerpt example (produced in Study 4 of this thesis), addressing a task involving planning a trip to a music festival. The students practised speaking in the system independently, which is not a new idea, as shown in the section giving a brief background to computer-assisted language learning (CALL) after the next section about language learning and the development of speaking skills.

## 1.1 SECOND-LANGUAGE TEACHING AND LEARNING – SPEAKING

In this thesis, the term second language (L2) is used as a shorthand (Goh & Burns, 2012; Chapelle & Sauro, 2017), referring to both English as L2, the language learnt after the first language or another dominant language, or to a third language (L3) such as French, German, or Spanish, so-called modern languages (National Agency for Education, 2018a). The term “target language” (Goh & Burns, 2012, p. 16) is used interchangeably with L2, referring to another language studied in addition to the student’s first language. Learning (and teaching) L2 are traditionally based upon four basic skills; two productive skills: speaking and writing, and two receptive skills: reading and listening (Li, 2017).

Speaking in L2 learning refers to the ability of a learner to communicate effectively in the target language through oral expression. Speaking is a crucial aspect of L2 learning, as it enables learners to engage in real-world communication and interact with others in the target language. Speaking is a “combinatorial skill” (Johnson, 1996, p. 155), and students must, simultaneously, not only have a knowledge of the language (grammar, vocabulary, pronunciation) and discourse, but also have the ability to produce grammatically correct utterances, using intelligible pronunciation, intonation, and words appropriate to the current social and cultural context. The student is also expected to understand and use a range of communicative strategies to compensate for any lack of language knowledge by using, for instance, interactional strategies such as asking the interlocutor to repeat a previous utterance in the dialogue. Hence, speaking is considered to be a challenging and complex learning process and, in Field’s (2011) words, “one of the most complex and demanding of all human operations” (p. 70), exerting heavy cognitive demands on students in real-time.

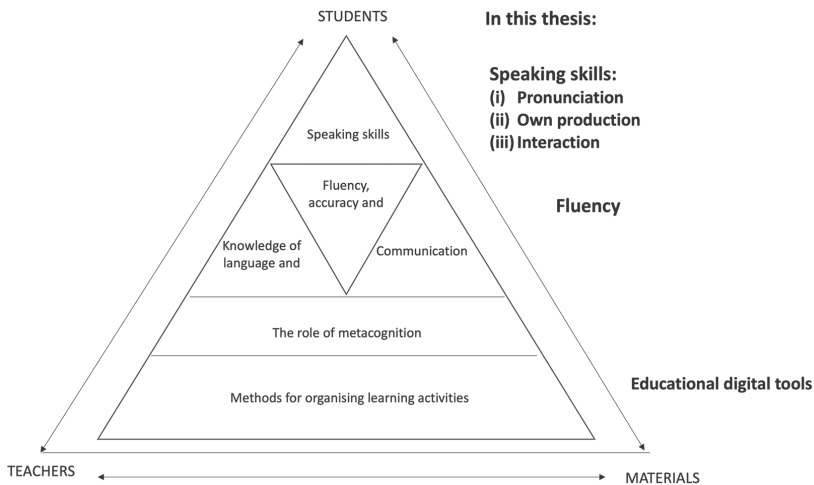
Engaging in meaningful language practice is believed to foster the development of second language (L2) speaking skills, as it involves actively using the language to convey the intended meaning. In the research literature, three features of speech with an impact on the quality of a student’s ability to perform are fluency, accuracy, and complexity (Goh & Burns, 2012; VanPatten & Benati, 2015). Fluency is represented in focus on the exchange of meaning, where the spoken message is communicated coherently and effectively. It can be measured by successful turn-taking (Li, 2017). Accuracy focuses on form, with the correct use of grammar,

vocabulary, and pronunciation. Complexity is a mixture of both fluency and accuracy, focusing on meaning and form in precise communication, developed into appropriate speech for the social and cultural context (Goh & Burns, 2012; Li, 2017).

Learning and developing L2 speaking skills can be conceptualised as a dynamic process that involves various components of speaking competence. To visualise speaking skills, and the components included, see Figure 1.2, which is based on Goh & Burns’ holistic approach to teaching speaking in their methodological framework (2012, p. 139). Three successful factors are the teachers, the students, and the materials. Their approach is used to frame and highlight the focus of this thesis, speaking skills, and what the concept of speaking skills is referring to when introducing SDS as a new method for teaching and practising L2 speaking for learning and development.

**Figure 1.2**

*Components in the teaching of speaking*



*Note.* Knowledge of language refers to having sufficient knowledge linked to speech production, grammar, phonology (sound system), and vocabulary. Knowledge of discourse refers to emphasis and tone. Adapted from Goh and Burns (2012, p. 139).

In this thesis, the speaking skills under study draw on the learning situations in the SDSs (Studies 2–4) and concern: (i) pronunciation (of words, phrases, and sentences), (ii) the student’s own production (for example, asking/answering questions), and (iii) interaction through dialogue and conversation (used interchangeably). The central subskill of speaking, interaction, naturally also involves listening skills (Hodges et al., 2012), the ability to listen and understand the other interlocutor while simultaneously producing one’s utterances in a comprehensible way to maintain the dialogue’s fluency, and still using adequate grammar and vocabulary for the social and cultural context. As expressed by Li (2017, p. 49): “the interpretations of other’s [sic] utterances and comprehension of ideas and the negotiation of meanings are carried out through interaction”. The focus is therefore mostly on fluency, but accuracy is also involved. Although students’ metacognitive awareness, gained through reflection upon their L2 learning and development can be seen as the foundation of teaching and learning to speak (Goh & Burns, 2012), metacognition knowledge is not the focus of the analysis in this thesis.

The ability to speak L2 is often difficult to “develop outside school and difficult to catch, methodologically, for the one who is going to define and characterise the student’s oral production and interaction” (Granfeldt et al., 2016, p. 2, my translation from Swedish). Spoken interaction is at its heart but this can be difficult to provide in a classroom context (Knapp et al., 2011; Timpe-Laughlin et al., 2020), and is also recognised as a challenge to provide in remote L2 education (Hartshorn & McMurry, 2020).

The complexity of spoken interaction and its transient nature makes it difficult to assess, and this, combined with physical constraints in the classroom, such as large class sizes, makes it challenging to give individual feedback (Goh & Burns, 2012). Hence, speaking is often neglected in favour of writing (*ibid.*). Students often report a lack of sufficient speaking skills and low self-confidence, and a clear majority consider speaking as challenging to learn, yet still the main target for L2 learning (Bygate, 2011; Darancik, 2018; Harlow & Muyskens, 1994; Pakula, 2019; Garbati & Mady, 2015). This is highlighted as a speaking paradox since being able to speak a language is considered by both teachers and students “as the most salient aspect of communicative competence”, yet it frequently receives insufficient attention in language education (Fernández & Andersen, 2019, p. 50). Additionally, previous studies have indicated that female

students experience higher levels of anxiety about L2 speaking in the classroom (Granfeldt, 2019; Öztürk & Gürbüz, 2013).

Meaningful opportunities and engaging tasks for practising speaking about content that is relevant to the students are conditions for successful learning and development within the communicative approach (Ellis, et al., 2020; González-Lloret, 2017). Such practice prepares students for everyday-life communication outside an educational context by stressing the active use of the target language in tasks or problems to solve.

The research field of second language acquisition (SLA) highlights key principles for the effective learning and development of L2 or L3 (Ellis, 2005; Li, 2017; Nation, 2007). This thesis focuses on the following L2 key principles relevant to the purpose of the conducted studies: “authentic input”, “conscious noticing on form”, “opportunity for interaction”, “in-time individualised feedback”, “low affective filter”, and “an environment where language can be used” (Li, 2017, p. 28; see Section 4.4). Additionally, emotions are considered to play an important role in L2 learning and development (Swain, 2013) and link the information tailored to students’ interests and everyday lives (Ellis & Shintani, 2014).

## 1.2 COMPUTER-ASSISTED LANGUAGE LEARNING – SPEAKING

When it comes to enhancing the learning and development of speaking skills, there have been around one hundred years of educational technology following the phases or approaches within the research field of CALL (Otto, 2017). The “talking machine”, a phonograph and groundbreaking technology for its time, was introduced to students to aid in pronunciation and listening practice, and although “the silent verdict brought in by its general abandonment is that it is not worth the trouble it involves”, Clarke (1918, p. 117) still recognised the benefits of recordings in L2 teaching and recommended that teachers investigate it. Influenced by behaviourism (Skinner, 1968) and the audio-lingual method (Richards & Rodgers, 2001), language laboratory classrooms provided native-speaker voices in drill exercises for a repetitive focus on listening and speaking (pronunciation) before reading and writing. These labs were common during the 1950s and throughout the ‘60s and ‘70s. They were followed by the development of interactive games and simulations in the ‘80s, which stressed

learning through interactivity and authenticity by using the target language in communication (Otto, 2017). The introduction of microcomputers facilitated the development of listening comprehension skills, promoting film, audiotape, and videotape. Various multimedia CALL systems were developed, predating today's communicative approach, where social interaction is emphasised as a means of achieving the goal of language learning (Bardel, 2019; Loewen & Sato, 2018). This speaking practice may be facilitated by, among other things, emerging conversational AI.

This thesis is positioned within the field of CALL, and specifically within the emerging subfield of dialogue-based CALL (DB-CALL; Bibauw et al., 2019), focusing on the learning and development of L2 speaking skills. Findings about the new educational landscape of L2 teaching and learning (Clark et al., 2016), the changing roles when introducing conversational AI, and its possibilities and challenges at the group and individual levels would benefit both teachers and school management. Such findings would also provide the DB-CALL field, with its interest in human-machine interactions and the development of SDSs with ECAs for L2 education, with new insights by highlighting any underlying aspects that seem important for the educational experience of practising L2 speaking with conversational AI.

### 1.3 TEACHERS' VIEWS ON AND EXPERIENCES OF DIGITAL TOOLS

Teachers are a central part of the design and arrangement of language teaching, and seek to facilitate language learning opportunities for students, both with or without the assistance of digital tools (Blake, 2017). Anderson et al. (2001) propose three activities of teachers: "the design, facilitation, and direction of cognitive and social processes for the purpose of realizing personally meaningful and educationally worthwhile learning outcomes" (p. 5).

However, the question of how to integrate and use digital tools in language teaching is both an urgent and complex issue (Li, 2017). Students must be supported by well-informed teachers and appropriate teaching materials (Goh & Burns, 2012). In Sweden, teachers have autonomy regarding the selection of teaching materials and their choice of digital tools, and how and when to use them, influencing the teaching space and

what teaching will be possible to conduct (Sofkova Hashemi & Cederlund, 2017). Furthermore, teachers' knowledge, pedagogical beliefs, and goals for teaching must coincide with the possible activities provided by the selected digital tools in order to be put into practice (Dimitriadis, 2012).

In comparison to reading, writing, and listening, teachers have ranked speaking as the most crucial skill to learn (Thiriau, 2017), but also as problematic to teach (Burns, 2016; Bygate, 2011; Goh & Burns, 2012), especially oral interaction and how to keep a conversation going (Erickson et al., 2018). As a springboard for the suggestion made in this thesis that conversational AI be employed to enable speaking practice opportunities for students in the target language, it therefore seemed logical to initially investigate whether Swedish language teachers found it feasible and were willing and able to apply it in an institutional, educational context.

In this thesis, digital tools refer to any educational technology, software, or application used in language education, including everything from digital teaching materials to digital resources on the internet, such as games, apps, or systems designed to support and enhance L2 teaching and learning. These could be interactive and/or multimedia resources involving a collaboration between the student's L2 learning and the tool for practising L2, such as vocabulary, grammar, pronunciation, listening or more holistic proficiency development. In general, digital tools can also enable various instructional techniques; for instance, feedback, progress tracking, or individually adapting the difficulty level and content to the student automatically in line with their performance and levels achieved (Utterberg Modén et al., 2023). These educational digital tools can incorporate spoken or written interaction in the form of teacher-to-student, student-to-student, or student-to-CA/ECA.

## 1.4 STUDENTS' EXPERIENCES OF CONVERSATIONAL AI

The main part (Studies 2–4) of this thesis took students' perspectives on practising language speaking skills with ECAs in selected systems by exploring the students' self-reported experiences (Levy, 2015). Participating Swedish lower-secondary students reported their experiences in terms of thoughts, views, feelings, and behavioural aspects connected to learn-

ing (Roth & Jornet, 2014) while practising speaking. The practice was intended to support their learning and the development of speaking skills through student–ECA interaction in the target language. Experience is considered crucial for learning (Roth & Jornet, 2014), and this motivated my choice of experience as the unit of analysis in relation to students’ speaking practice.

Given the novelty of using this type of system in an actual educational context with new forms of interactional learning activity, a framework for student–CA interaction (FoSCAI) was iteratively developed. This enabled the operationalisation of the experience of practising speaking the target language in an SDS (Ericsson & Jansson, 2021). The experience of practising was hence dimensionalised into cognitive, emotional, social, and teaching aspects, which were studied and analysed both separately and collapsed together, as suggested in FoSCAI. The teaching dimension refers to the feedback and supportive features of the system, which are provided to set the level of challenge in the students’ practice. Since speaking with an ECA is the central activity in the practice, the students reported separately on their speaking experiences (involving the same dimensions as above, apart from teaching), which were then also included in the educational experience. Semiotic resources, such as non-verbal conduct in the ECA’s social cues e.g., facial expressions, gestures, etc. (McTear, 2020), were included in the experiences of the student-to-ECA interaction that were explored.

As proposed by Reeves and Nass (1996), people tend to treat human-like entities as if they were real human beings, interacting socially and emotionally with them and having the same expectations of the interaction as they would of an interaction with a human being. Therefore, it is valuable to explore the students’ self-reported experiences of speaking with ECAs, to determine if, and if so how they related socially to them in the interaction and what feelings the students developed towards the ECAs. Additionally, it is interesting to explore what feelings and signals were ascribed to the ECAs themselves, a tendency referred to in the literature as humanisation, or anthropomorphisation (*ibid*). The educational experience is hence a comprehensive experience of all four dimensions together, including the central speaking experience. When students rate it themselves, the educational experience is referred to as the overall experience.



## 1.5 AIM AND RESEARCH QUESTIONS

The aim of this thesis is to explore and gain a nuanced, critical, and comprehensive understanding of how conversational AI is experienced by students during their institutional language education when practising speaking skills through face-to-face interaction. Based on this aim, three research questions (RQs) were formulated.

RQ 1: Focusing on speaking skills, what are language teachers' views and experiences of teaching and using digital tools?

RQ 2: How do lower-secondary students experience practising L2 speaking skills in an SDS?

RQ 3: How do lower-secondary students experience speaking L2 with an ECA?

As displayed in Table 1.1, Paper I reports on a teacher study (Study 1), addressing RQ 1, and Papers II–V report on three student studies (Studies 2–4), addressing RQs 2 and 3. For details and elaboration of how the RQs are addressed in each paper, see Chapter 7.

**Table 1.1**

*Research questions in relation to the papers and studies of this thesis*

<b>Research Question</b>	<b>Paper</b>	<b>Study</b>
RQ1	Paper I	Study 1 (with teachers)
RQ2	Papers II–V	Studies 2–4 (with students)
RQ3	Papers II–V	Studies 2–4 (with students)

## 1.6 OUTLINE OF THIS THESIS

This thesis is divided into two parts: (i) a text jacket consisting of ten chapters wrapping the papers, and (ii) the collection of papers (Papers I–V), reporting on Studies 1–4. The structure of the first part is as follows:

*Chapter 1. Introduction* gives a brief background to the research interest, and motivates and presents the aim and the RQs.

*Chapter 2. Language Education in Sweden* outlines the empirical context of this thesis.

*Chapter 3. Conversational AI* outlines the technology investigated in this thesis, SDSs and ECAs.

*Chapter 4. Dialogue-Based Computer-Assisted Language Learning - Speaking Skills* first introduces the rationale of this emerging field, within which this thesis is situated. Then follows a brief historical overview of its interrelated fields, SLA and CALL, to situate the SDSs and their theories of learning. The concept of speaking skills is introduced, alongside the ideas of cognition and emotional factors and some established key principles for effective L2 learning and development.

*Chapter 5. Related Research* covers previous studies of teachers' views and experiences of teaching speaking skills and the use of digital tools, as well as students' experiences of practising L2 speaking skills in SDSs and speaking experiences with (E)CAs.

*Chapter 6. Theoretical Underpinnings* first provides an introduction to this thesis' central theories, framing the understanding of language learning in communication, social interaction and the Interaction Hypothesis in relation to active learning and the development of L2 speaking. Then the concept of experience is presented. Finally, the chapter describes the developed methodological and analytical framework FoSCAI as a refinement and conceptualisation of the educational experience in an SDS.

*Chapter 7. Research Design and Methods* describes the research design and multiple methods used to address the aim and RQs of this thesis.

*Chapter 8. Summary of the Pilot and Papers* gives a summary on a paper-by-paper basis in relation to the three RQs.

*Chapter 9. Contributions and Discussion* discusses the key findings and knowledge contributions aligned with the three overarching RQs addressing both teachers (RQ1) and students (RQs 2-3). Moreover, this chapter provides methodological contributions, methodological reflections, and practical implications, before concluding with suggestions for future research.

*Chapter 10. Conclusion* summarises how the RQs were addressed and answered.

*Chapter 11. The Swedish Summary* gives an overview of this thesis in Swedish.

## CHAPTER 2

# LANGUAGE EDUCATION IN SWEDEN

This chapter presents a brief description of language education in Swedish lower-secondary schools, emphasising the study of languages in compulsory education - English, and modern languages - and the communicative and action-oriented approach in relation to speaking skills. The first section provides a brief overview of the national curriculum (National Agency for Education, 2022b; 2018a, English version) and syllabus for grades 7–9, English and modern languages, which reads similarly (National Agency for Education, 2022c, 2022d) and is tightly linked to the Common European Framework of Reference for Languages (CEFR; Council of Europe, 2020, 2022). To understand the theoretical underpinnings of the current, action-oriented approach to language teaching, the most central SLA constructs are provided. Finally, the chapter ends with a presentation of digitalisation in Swedish language education.

### 2.1 STUDYING LANGUAGES

In Sweden, school starts at the age of 6. English (as L2 for many students) starts as a compulsory subject at some point during primary school, depending on how schools distribute the subject's total hours. It is compulsory at the latest from grade 4, and continues through lower secondary until grade 9. At the age of 12, children also choose a second European language, often referred to as L3, German, French or Spanish, which in Swedish curricula are labelled modern languages (*moderna språk*, in Swedish; National Agency for Education, 2018). It can be studied as the “Language choice” (*Språkval*, in Swedish) or as the subject named “Student's

choice” (*Elevens val*, in Swedish.; National Agency for Education 2018; 2022d). The English and modern languages syllabuses are very similar regarding content, learning objectives, and performance standards, but the progress is faster in English, with higher-level demands.

Approximately 85% of students in grade 7 choose to study a modern language (L3), and half of these choose Spanish (National Agency for Education, 2020). Approximately 65% of students who graduate from lower secondary school have a qualification in modern languages (*ibid.*), and all modern language courses struggle with student dropout (Swedish School Inspectorate, 2022; Tholin, 2019). Modern languages are not compulsory in lower-secondary school, and students can take another subject instead, additional Swedish or English.

Teaching in modern languages often lacks individually adapted teaching (Swedish School Inspectorate, 2022). The Swedish School Inspectorate’s latest report shows that schools have not sufficiently analysed the dropout rate for modern languages in relation to the quality of education, and that many students dropped out due to a lack of support. Furthermore, it reports that students often lack the opportunity to practise communicating in the target language in authentic, everyday-life situations, especially spontaneous speech. However, many teachers were reported to have created inclusive teaching, where it was ok to make mistakes. The European Union Parliament (2017) aims for all citizens to master two languages (key competencies) as well as their first language, and stipulates that they must start learning them early in life (European Commission 2019a, 2019b), which marks out the high Swedish dropout rates for modern languages as a partial failure. In the latest report from the Eurydice Network on language teaching in Europe, Sweden was reported as being equally as good as, or better than, other European countries that do not have two official languages (European Commission, 2023). Overall, 79% of secondary students are studying two foreign languages in Sweden (versus an average of 60% in Europe as a whole). To help the status of modern languages in Sweden and increase the number of students who choose a modern language and continue to study it over time, grade-point average enhancement credits (*meritpoäng*, in Swedish) are used, which raises the student’s general grade-point average above the basic level (Granfeldt et al., 2021).

### 2.1.1 MODERN LANGUAGES AND ENGLISH

There have been various national efforts over the last decade to encourage students to learn languages and to complete their courses in modern languages, but the desired outcomes have not been achieved (Tholin, 2019). Three interacting explanations have been suggested: no existing central school administration for the concretisation of the political intentions, teachers not being aware of the importance of implementing the reforms, and a gap between the actual underlying factors influencing students' lack of motivation and interest in studying modern languages. Students' lack of motivation and interest is explained by modern languages being too demanding and of limited use in comparison to English and students feeling that the teaching is boring.

However, students participating in the project *Learning, Teaching, and Assessment of Second Foreign Languages: An Alignment Study on Oral Language Proficiency in the Swedish School Context* (TAL, original italics) generally reported positive attitudes towards learning another language as well as English, but only 40% of the students found modern languages to be a school subject of importance (Erickson et al., 2022, p. 159; TAL, n.d.). There was an identified need to develop spoken interaction and spontaneous speech, and the students expressed anxiety about speaking the target language in the classroom, girls somewhat more so than boys (Granfeldt, 2019). Aronsson (2020), in line with earlier results, showed in her studies that Swedish students of Spanish do not practise speaking enough in school because they mostly focus on writing, and generally do not reach the expected A2 level (Council of Europe, 2020) by the end of lower-secondary school in terms of speaking. Additionally, based on students' self-reports, she found that they estimate most spoken activities to consist of reading vocabulary aloud or preparing written manuscripts, focusing on form. Aronsson (2023) revealed that speaking is the least frequent learning activity in Spanish classes, also with a clear dominance of form training, which is not in alignment with the communicative, action-oriented approach in CEFR. Since out-of-class exposure to the target language is limited for modern languages in Sweden, the impact of the design of in-class learning to provide meaningful learning opportunities is crucial. There are still only limited studies on the teaching and learning of modern languages in Sweden (Swedish Research Council, 2019; Erickson et al., 2022).

Learning to speak English, on the other hand, is somewhat different in Sweden due to its special social status, and it is “unofficially taking the role of a second language in Swedish society” (Bardel et al., 2019, p. 12). English is generally heard more often in the media and the linguistic landscape in Sweden than French, Spanish, or German, which consequently are studied in an input-poor context. Already at an early age, extramural English comes naturally more often to students outside school through TV, radio, music, and Internet activities enabling spoken interaction through online video games. Sitting at the computer, young male students are found to use their English more in games than females (Sundqvist & Sylvén, 2014). English functions as a *lingua franca*, both in Europe and globally, and is used as a common language for communication by speakers who do not have the same first language (Mitchell et al., 2012). However, in Sweden, the official language is Swedish, along with five minority languages (Government Office, n.d.), which underlines the importance of enabling opportunities for speaking practice in L2 language education.

In this thesis, L2 refers to any target language (English, French, German, or Spanish) being taught or learnt in an institutional educational context, while retaining an awareness of the distinct social context of learning these languages in Sweden, and the pedagogical differences. The focus of this research is on experiences of using conversational AI to practise speaking skills in any target language by interacting with ECAs. Target language refers to the language that the students are striving to learn and develop. The following section presents how speaking is addressed in some national and European policy documents.

### 2.1.2 COMMUNICATIVE AND ACTION-ORIENTED APPROACH

In Sweden today, there is a preference for a communicative and action-oriented approach to teaching, learning, and assessing languages (Bardel, 2019). The emphasis is on developing students’ implicit knowledge, which will enable increased communicative ability in various oral and written contexts and provide them with opportunities to interact with authentic language and culturally authentic materials, based on the assumption that effective language learning mainly occurs through interaction. This com-

municative approach involves spoken interaction as both the means and the goal of language learning. The Interaction Hypothesis (Long, 1996) is the primary theoretical model for how this can be applied in a social context (see further details in Chapter 6).

The updated national curriculum specifies that language teaching must aim to develop students' ability to "understand spoken and written language, being able to produce [one's] own utterances and interact with others orally and in writing and adapt one's language to different situations, aims and interlocutors" (National Agency for Education, 2022c, English, my translation from Swedish). The Swedish curriculum also highlights core content (*centralt innehåll*, in Swedish), such as "everyday-life situations, interests, people, places, activities, and events" (National Agency for Education, 2022d, modern languages, my translation from Swedish) that is relevant to the students. Communicative ability includes "linguistic certainty and using different strategies to facilitate communication when one's language skills do not suffice" (National Agency for Education, 2022d, modern languages, my translation from Swedish). To implement these goals, the Swedish approach to teaching speaking skills can be interpreted as emphasising:

- (i) Communicative ability: Teaching must focus on developing students' ability to use the language for meaningful communication, rather than memorising vocabulary and grammar without context (Ellis et al., 2009).
- (ii) Task-based language teaching (TBLT): A student-centred approach emphasising practising L2 in authentic tasks that are meaningful and relevant to the student, promoting learning through use of the target language in everyday-life situations and contexts, such as role-plays, discussions, and presentations (Ellis et al., 2020; Long, 2015).
- (iii) Interactivity: teaching must provide opportunities for students to interact with each other and with authentic materials such as films, songs, and texts.
- (iv) Intercultural competence: Teaching must help students to develop an understanding of the cultural contexts in which the language is used in order to encourage respect for different cultural perspectives.

This action-oriented approach emphasises active and autonomous learners *using* the language, rather than it just *being a subject* for study in school. When implementing the action-oriented approach:

The methodological message of the CEFR is that language learning should enable learners to act in real-life situations, expressing themselves and accomplishing tasks of different natures. Thus, the criterion suggested for assessment is communicative ability in real life, in relation to a continuum of ability (Levels A1–C2). (Council of Europe, 2020, p. 29)

Communicative competence, including linguistic, sociolinguistic, and pragmatic competence, is the goal of language education. Communicative activities, such as conversation or information exchange, are interactive and signal that “the participants alternate as producers and receivers, often with several turns” (Council of Europe, 2020, p. 58). Language teaching and learning are organised into six proficiency levels, referring to students’ overall ability, ranging from basic language to the proficient, independent user (A1–C2; *ibid.*, p. 36). The levels are illustrated with can-do descriptors related to everyday-life situations where the students are involved in the learning process (*ibid.*, p. 35).

Swedish schools employ five levels of language grades, with students receiving a passing grade of E to A. For the highest grade A in (i) modern languages and (ii) English in the last year of lower-secondary school, the criteria are:

- (i) The student listens, understands, and interprets **the main content and essential details** in clearly spoken, simple language at a slow pace about everyday life and familiar topics. (...). In spoken production of various kinds, the student expresses him- or herself **to some extent with variation, relatively clearly and coherently** (...). In interaction, the student expresses him- or herself **relatively clearly using words, phrases, and sentences**, and **to some extent adapted** to the aim, receiver, and situation. In addition, the student chooses and uses strategies that facilitate and improve the interaction. (National Agency for Education, 2022d, my translation, original bold).



- (ii) The student listens, understands, and interprets in a **well-grounded** and **nuanced way as well as overall** and details in spoken language at an average pace and in different contexts. (...). In spoken production of various kinds, the student expresses him- or herself with some variation, clearly and coherently. The student also formulates him- or herself **with fluency** and with **certain adaptations** to the aim, receiver, and situation. (...). In interaction in different contexts, the student expresses him- or herself **clearly and fluently** and **with certain adaptations** to the receiver and situation. Additionally, the student uses strategies that facilitate and improve the interaction and **help it to advance constructively**. (National Agency for Education, 2022c, my translation from Swedish, original bold).

#### 2.1.2.1 EDUCATIONAL ACTIVITIES FOR L2 SPEAKING PRACTICE

The teacher is expected to create possibilities and activities for students to learn and develop skills, with the support of pedagogical material, sometimes including the assistance of technology (Li, 2017). It is challenging to provide meaningful, authentic opportunities for students to practise speaking in the target language in everyday-life scenarios (Bibauw et al., 2019; Goh & Burns, 2012; Knapp et al., 2011). Speaking activities can be created in various ways, both synchronously and asynchronously, through arrangements for learning both inside the institutional context and in naturally occurring contexts outside institutional environments (Bygate, 2011; Goh, 2017). However, there is no disagreement that immersion in the target country is the ideal context for language learning, where all the key features of speaking can be naturally stimulated in spoken interaction with native speakers (Sieloff Magnan & Lafford, 2012).

Linguistic input is important, thus opportunities should be provided for dialogues preparing the student to cope with everyday-life situations in the target language (Littlewood, 1984; 2018). Typically, one way is to organise students into small groups “to develop learners’ speaking abilities (...) where they can practise communicating in the target language” (Goh, 2017, p. 247). According to Bygate (2011), all the key steps of language processing have to be taken into consideration when planning

teaching activities involving speaking skills in order “to enable learners to engage with the social and psychological processes of speech production, in real-time” (p. 410). These can include both form-focused and more meaning-focused exercises and be structured so as to elicit prepared language use, or more “authentic activities involving creative and spontaneous speech production” (Littlewood, 2004, p. 322). There are various ways in which students can actively engage in spontaneous and meaning-focused L2 speaking activities through spoken interaction, giving them a social context (Littlewood 1984, 2018; Bygate, 2011) for L2 development; for instance:

- (i) Spoken interaction, face-to-face dialogues in classroom settings, with students in groups, pairs of two students, or teacher–student, in spoken exercises, simulated situations in role-play activities, and dialogues (Goh & Burns, 2012; Littlewood, 2018).
- (ii) Spoken interaction face-to-face through synchronous conversation at a distance by mediating technological tools, such as Skype or Zoom, or by playing interactive games (Walker & White, 2017).
- (iii) Spoken interaction with conversational AI (synchronously or asynchronously) through maintaining a dialogue with a CA to solve a task in an everyday-life scenario, for instance, ordering in a café (e.g., Bibauw et al., 2019).

Today, there exists a plethora of digital tools for language education, including a few systems that apply conversational AI for speaking practice in DB-CALL. The necessary devices and infrastructure for applying them are mostly already set up in the Swedish school context. The next section describes digitalisation in Swedish education.

## 2.2 DIGITALISATION IN SWEDISH (LANGUAGE) EDUCATION

Overall, Swedish education is well equipped with digital devices and digital tools. Many schools have one-to-one programmes, meaning that all students and teachers have access to one digital device each, such as an iPad, Chromebook, or computer (Tallvid, 2015). Digitalisation has led to significant changes in how education may be delivered, but these opportunities also come with challenges (National Agency for Education, 2018b). It includes the transformation of the roles of teachers and students and in the ways in which they interact with each other, and also with the digital tools that are available to support and enhance teaching and learning (Bates, 2019; National Agency for Education, 2018; Utterberg Modén et al., 2023). However, the Swedish school system as an institution has a long history connected to its physical environment, with the classroom setting being pretty much the same today as it was a hundred years ago, and it is sometimes difficult to make good use of emerging technologies (Säljö, 2022).

The curriculum for compulsory school was rewritten in 2018 to highlight adequate digital competence and strengthen the use of digital tools in all school subjects at all levels (National Agency for Education, 2018a), and was further revised in 2022 (National Agency for Education, 2022b). The Swedish government's first general national digitalisation strategy was introduced in May 2017 (Government Office, 2017) and its strategy for the use of AI in 2018 (Government Office, 2018), in alignment with the overarching national goal to be "the best in the world at making use of the possibilities of digitalisation" (Government Office, 2017, p. 10, my translation from Swedish). In education, it was followed up in the Action Plan for school digitalisation (Skoldigiplan, 2018). The digitalisation of language education is regulated by the compulsory school curriculum, which stresses that all pupils should be given the opportunity to develop their ability to use digital technology (National Agency for Education, 2018a). However, it was not explicitly mentioned in the revised syllabus for the specific subjects modern language or English (National Agency for Education, 2022c, 2022d).

A suggestion for a new version of the Swedish National Digitalisation Strategy for the school system for 2023–2027 was presented in late 2022, further strengthening the need for digital competence and the aim of using digital tools in education to deliver better quality and opportunities for lifelong learning and improve quality (National Agency for Education, 2022a). According to this strategy, there should be good and equal access to digital tools, and the possibilities of technology should be utilised effectively in order to develop education (p. 12). However, the Swedish government announced a pause of the planned strategy in early spring 2023 (Letmark, 2023), and the schools minister compared the digitalisation strategy with an experiment (Edholm, 2022). The political and scholarly debate is ongoing at the time of writing this thesis (e.g., Beijmo, 2023; Thorell et al., 2023) and signals a shift towards promoting more traditional forms of education (Selwyn, 2023b).

This chapter has outlined the Swedish educational context within which this thesis is set, which is framed by central aspects of the Swedish and European policies setting out the communicative approach of languages and digitalisation. The technicalisation of Swedish L2 language education over time is not treated separately, but is considered to follow fairly similar patterns, and to be presented in the included overview of CALL and DB-CALL (see Section 1.2 and Chapter 4). The next chapter presents the technology introduced in this thesis: conversational AI.

## CHAPTER 3

# CONVERSATIONAL AI

This chapter describes conversational AI as used in this thesis to enable an understanding of the technology and terminology behind the human-machine interaction taking place in the SDSs used, including virtual agents such as ECAs (Cassell & Bickmore, 2003; McTear, 2020). Then the two types of ECAs in this thesis are outlined. Finally, conversational AI is briefly put into a historical perspective, ending with its use in education today, and possible future uses.

### 3.1 SPOKEN DIALOGUE SYSTEMS AND VIRTUAL AGENTS

The computer-based interface of SDSs within DB-CALL enables conversation in meaning-focused interactions for students to practise their L2 speaking skills (Bibauw et al., 2019). These systems may use predefined dialogues or dynamic ones, depending on the user's input. Morton and Jack (2005) have suggested using Long's interaction hypothesis (1996) as a basis for designing SDSs, thus enabling input, output, feedback, and the negotiation of meaning (see Chapter 6 for more about the theoretical underpinnings).

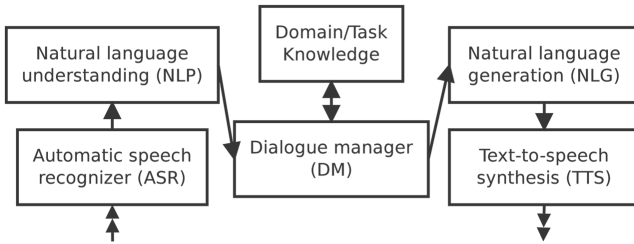
Technically, a typical AI-empowered SDS involves five major functionalities to handle the user's spoken utterances: automatic speech recognition (ASR), natural language understanding or processing (NLP), dialogue management, natural language generation, and text-to-speech synthesis (Griol & Callejas, 2016). Figure 3.1 illustrates the basic architecture of an

SDS, including these major functionalities. There are sets of basic actions that the systems have to undertake cyclically and systematically after each user utterance within the system:

- Recognise the user's words (ASR)
- Interpret and understand the useful information from the words and utterances (NLP)
- Adapt the interaction to its context (DM)
- Access web services and the system to locate information
- Decide about suitable actions after the user's utterance (DM)
- Choose a suitable utterance and respond to the user (NLG, TTS)

**Figure 3.1**

*The basic architecture of an SDS*



*Note.* The user's input as speech initially comes into the system's ASR, illustrated with the upward-pointing double arrow. The spoken form of the system's response is produced in the TTS and is illustrated by the downward-pointing double arrow (Alès et al., 2012, p. IV).

In an SDS interface, there are virtual agents or voice-based chatbots. These are designed to simulate human conversation and interact with users in natural language through their voice by interpreting a user's spoken input and generating a verbal response. There is, however, confusion around the terminology, with no less than 161 synonyms found for describing conversational AI systems (Walker, 2023). These terms vary between interrelated research fields, which complicates meta-analyses and the comparison of

results between different studies. This chapter highlights some terms of importance for this thesis.

Virtual agents are often labelled depending on the agent's function and purposes, such as conversational agents (CAs) or embodied CAs (ECAs), which are designed to simulate conversation and solve given tasks (McTear, 2020; Bickmore & Cassell, 2005), and pedagogical agents that lecture in a one-way fashion (Johnson & Lester, 2016). Pedagogical agents are designed to support learning and teaching activities (tutoring) while providing feedback and guidance; no reply is necessarily expected from the user. They are mostly designed as animated characters for tutoring systems (Graesser et al., 2017), instructional videos (Craig & Schroeder, 2017), and educational games (Wang, 2020) in different learning domains; for instance, healthcare training (Albright et al., 2018) and virtual receptionists (Chilufy & Arvola, 2021). Since there is a possible overlap between these research fields, results from pedagogical agent studies are also considered in this thesis.

ECAs are often designed as animated, human-like characters to simulate a physical presence and establish human-computer relationship (Bickmore & Rosalind, 2005). The ECAs express body language, facial expressions, conversational gestures, and other nonverbal behaviours in order to simulate human appearance and behaviour as closely as possible in a conversation (McTear, 2020; Bickmore & Cassell, 2005). Studies within education have shown that using an ECA can enhance user engagement and satisfaction, and that exploring their social cues is important (Burden & Savin-Baden, 2019). Within L2 education, ECAs have been found to be effective as trainers in areas such as vocabulary learning, where the agents' iconic gestures matching human gestures have been shown to be beneficial for students' memory performance (Bergmann & Macedonia, 2013).

In parallel, some scholars use the term virtual humans for human-like embodied characters such as ECAs (e.g., Burden & Savin-Baden, 2019; Johnson, 2019a). This term was initially used by Johnson and Lester (2016), as well as Craig and Schroeder (2017), when referring to animated characters as pedagogical agents. Virtual humans, referring to characters with similar inherent functions as ECAs, are also used in various domains within education, such as trainee teacher education with virtual students (Howell & Mikeska, 2021), for practising communication skills with vir-

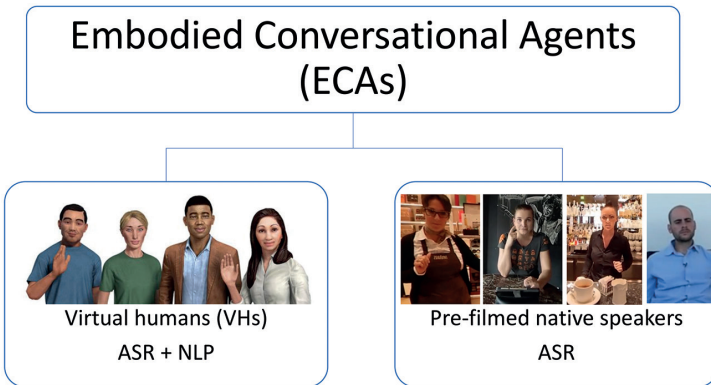
tual patients in simulated clinical learning situations to support speech pathology education (Robinson et al., 2020), health workers communication with virtual patients, (Johnson, 2021) and language learning (Johnson, 2019).

### 3.1.1 EMBODIED CONVERSATIONAL AGENTS

In this thesis, two types of ECAs were used in the selected SDSs, as illustrated in Figure 3.2. They are designed and programmed with the goal of interacting with a student in natural language, being their “communicational counterpart” (Bibauw et al., 2019, p. 7). The system’s users, here the students in Studies 2–4, are hence communicating *with* the computer, in comparison to *through* the computer in other computer-mediated forms of communication in the target language (e.g., video conferencing systems such as Skype or Zoom).

**Figure 3.2**

*Examples of embodied conversational agents employed in this thesis*



*Note.* © Alelo Inc and © ImmerseMe. Used with permission.

The visually represented interactive human-like ECAs, either male or female, were of two types. They are distinguished in this thesis by using the terms (i) virtual humans for the animated ECAs interacting in a vir-



tual environment in the SDS Enskill (Alelo, 2022), and (ii) pre-filmed human beings for the native speakers of the target language interacting in a 360-degree virtual environment in authentic settings in the SDS ImmerseMe (n.d.). The role of these two types of ECA was to facilitate opportunities for practising L2 speaking skills, which is part of the functionality offered in SDSs. The various constraints are predetermined in order to balance the conversational freedom; consequently, the agent's capacity to express itself orally and to interpret the student's spoken utterances affected how the interaction was carried through and the level of complexity that was attainable, as explained in more detail in Section 7.1.4.3 about the two selected SDSs. In order to situate the conversational AI used in this thesis in a historical context in these fluid times of emerging technology, the next section gives a brief overview of the history of conversational AI.

### 3.2 PAST, PRESENT, AND FUTURE

Although Pieraccini (2012) has pointed out that the evolution of speech in a machine can be traced back to the early 18th century, one of the earliest recognised modern NLP programs giving an illusion of understanding natural language was devised in the 1960s. This was the chatbot ELIZA (Weizenbaum, 1966). Through a set of pattern matching, it provided generic responses in human-like interactions, enabling simulated conversations between human and machine, imitating a psychotherapist in text-based dialogue systems.

Furthermore, Winograd's successful SHRDLU (Winograd, 1972), featuring block worlds and limited vocabulary for interaction in English, "demonstrated that for a carefully constrained dialogue in a limited domain it was possible to deal with meaning in a fairly comprehensive way, and to achieve apparently natural communication" (p. 214), although some problems with the program remained to be resolved (see further, Winograd, 1984). In their reaffirmation, Dreyfus and Dreyfus (1988) presented their critical standpoint on the status of the computers and AI of their time during the 1980s. Problematising natural language, self-awareness, and raising a warning flag for AI in education, they argued that machines were limited in their ability to replicate human cognitive processes and interact like humans.

ASR's first breakthrough in real speech was in the late 1980s, followed by conversational speech development during the '90s and early new millennium. Around 2014, there was a breakthrough in machine learning, which was used in virtual assistants and chatbots, enabling them to interpret speech into words in order to understand and respond to user requests, typically through voice or text interactions (Liu & Forsythe, 2021). Those assistants are examples of SDSs, and can be used for many tasks, from controlling smart-home devices to answering questions and from providing information to booking appointments. Today, digital voice assistants are common in some parts of the world, and in the USA, one-third of the population owns a smart speaker (Edison Research, 2021, March 11). In Sweden, 5% of the population report having a smart speaker (Swedish Internet Foundation, 2019) at home; for instance, a Google Assistant (n.d.), which is example of voice-directed systems using ASR and NLP for contextual conversations, enabling a virtual assistant to do things for you at home. Another AI-based interaction function from Google is Duplex (2018), which can call a restaurant and book a table for you (Leviathan & Matias, 2018). SDSs are also used, for example, in automated virtual assistants online in various social contexts, such as in smartphones (Apple's Siri, Samsung's Bixby, etc.), which are owned by 75% of Americans (Edison Research, 2021) and by 92% of Swedes (Swedish Internet Foundation, 2019) and claimed to be a part of our social interaction (Weilenmann, 2021). Additionally, virtual assistants are used in telephone-based services and transactions, travel inquiries, and reservations for flights or restaurants, sometimes on the phone, or in a chat online, with chatbots or an embodied character. These can consist of either verbal or written interaction, or a mixture of both.

### 3.2.1 APPLICATION IN LANGUAGE EDUCATION

Springing out of telephone interfaces for customer services, SDSs in language education have developed more multimodal interface systems since the 1970s. Over the last ten years, these have enabled synchronous computer-mediated communication, during which students practise speaking in the target language in task-oriented conversations *with* ECAs (Bibauw et al., 2019). SDSs use spoken and written interactions (transcripts of spoken dialogue) and other visual information via the ECA in the system's

animated or 360-degree video environment for the contextualised simulated dialogue. SDSs constitute a virtual learning environment that can be immersive to varied extents, using head-mounted equipment for high-immersive VR, or any digital device that can be connected online and support the SDS for low-immersive VR (Kaplan & Gruber, 2019). However, conversational AI is rarely used in an institutional, L2 educational context (Zhai & Wibowo, 2023, and SDSs for education often remain at a prototype level (Bibauw et al.2019).

Some examples of publicly available SDSs with millions of users worldwide are the popular Duolingo, used for pronunciation and speaking practice (n.d.), Rosetta Stone for feedback on pronunciation (2023), Babbel, which focuses on pronunciation and accuracy of speaking (2023), and Mondly (standard, for kids, VR, and AR) for daily language lessons around specific topics (2023). Another system that is available for English as a second language is ELSA Speak (English Language Speech Assistant) for pronunciation practice and adjustment (2023). However, in a review, the use of speech technology in these tools was found to provide insufficient feedback (Bajorek, 2017). The two internationally established SDSs employed in this thesis, Enskill (Aleo, 2022) and ImmerseMe (n.d.), were the only two encountered in 2019 that enabled real-time spoken interaction with ECAs in everyday-life scenarios. Today, Enskill has over 500 000 L2 student users in 25 countries and online, and ImmerseMe is licensed to 75 000 L2 students, primarily in New Zealand and Australia.

Another technology that is also of interest for language education was launched at the end of 2022 by Open AI in the trained powerful language model, Chat Generative Pre-Trained Transformer (ChatGPT; OpenAI, 2022), which generates texts based on the user's text input. Chat GPT has experienced rapid growth and has led to an ongoing engaged debate about conversational AI's benefits and challenges in education (Dilmegani, 2023). Friedman (2023) has claimed that we are now entering a new Promethean moment, where we have to harness the power of the rapidly evolving Generative AI, which presents us with both opportunities and challenges. It has been claimed that ChatGPT will have a forthcoming impact on L2 education by supporting conversational AI, and providing more personalised learning experiences and opportunities for encountering the target language (Cheong Hin Hong, 2023). According to Clark (2023) we will have a revolution for language learning. Enabling natural

oral interaction by adding voice to ChatGPT suggests the destiny of language education: “ChatGPT just killed the traditional language learning” (Registre, 2023). This rapidly emerging technology has the potential, in the future, to facilitate speaking practice conversations, and this generative AI has recently also been combined with the humanoid robot Ameca, which interacts unhindered in English, Japanese, German, Chinese, and French, and is even aware of variants such as British and American English (Evans, 2023; Engineered Arts, 2023). A pocket version is presented in the app Call Annie (2023; Sarwar, 2023), a female ECA that is able to hold a spontaneous conversation (see Figure 3.3), presenting herself like this:

Hey there! I'm Annie, your new best AI friend and (kind of) super intelligent assistant. Let me introduce you to CallAnnie, the app that allows us to connect in real-time through video calls, so you can talk to me with you [sic] wherever you go!

### Figure 3.3

*Pocket Version of an ECA in the app Call Annie*



*Note:* Maring, J. (2023)

In this thesis, the majority of the technology applied (Studies 2–4) is conversational AI, referring to SDSs with ECAs. For a further description of the selected SDSs and selection criteria set, see Chapter 7. These systems position this thesis within the subfield of DB-CALL, which is outlined alongside its related research fields in the next chapter.

## CHAPTER 4

# **DIALOGUE-BASED, COMPUTER-ASSISTED LANGUAGE LEARNING – SPEAKING SKILLS**

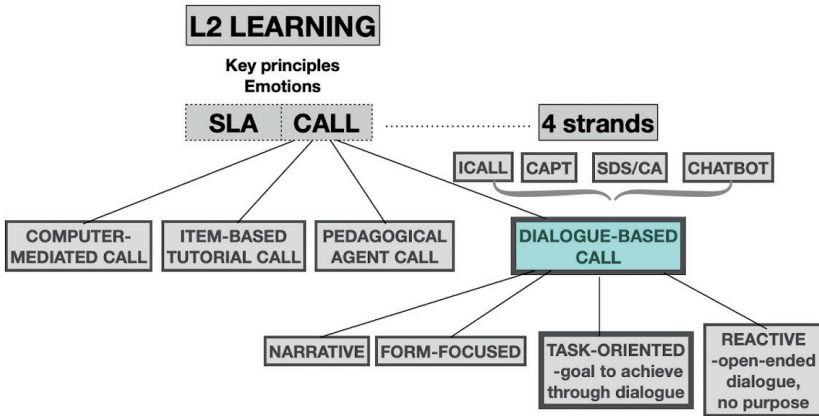
The field within which this thesis is situated, DB-CALL, is briefly presented in this chapter. There then follows an overview of the two closely related fields: L2 learning without technology, SLA, and L2 learning with technology, CALL. Finally, some key ideas and established principles for effective language learning that are central to this thesis are highlighted.

### 4.1 DIALOGUE-BASED, COMPUTER-ASSISTED LANGUAGE LEARNING

DB-CALL is an umbrella term for research that has emerged during approximately the last forty years within various fields and traditions. Bibauw et al. (2019) highlight that it is a dispersed and fragmented field that applies a wide variety of concepts, with implications from various disconnected research fields not yet making use of each other's findings. As illustrated in Figure 4.1, they identified four related research strands using different technologies for various focuses: (i) intelligent CALL (ICALL) for written modality, (ii) computer-assisted pronunciation training (CAPT), (iii) SDS/CA, for spoken modality, and (iv) chatbots, for written modality.

**Figure 4.1**

*Overview of DB-CALL (highlighted in colour), with its surrounding fields, subfields, four strands, and various approaches*



Bibauw et al. (2019) have proposed an operational definition of DB-CALL, which includes systems that allow students to interact in spoken or written form with an automated agent as interlocutor. This agent represents a native speaker of the target language and provides L2 practice, with feedback aiming to improve the student’s L2 skills. The unit of instruction is the pragmatic unit, the dialogue, where meaning is constructed through turn-taking between the student and the interlocutor, and solving a task is the goal of the simulated interaction. This definition distinguishes DB-CALL from the other identified subdomains of CALL, such as computer-mediated CALL, item-based tutorial CALL, and pedagogical agent CALL (see Figure 4.1).

DB-CALL sprang out of a developed typology based on a combination of the levels of constraints on meaning and form in the dialogues and systems, and Bibauw et al. (2019, 2022) have suggested a conceptual framework for classifying it based on the following criteria: (i) interactional implications, (ii) instructional characteristics, and (iii) technological features. According to Bibauw et al. (2022), there are four types of DB-CALL, within which most SDSs are classified as task-oriented systems,

where the student strives to achieve a goal by means of the dialogue (e.g., book a train ticket), or reactive systems, with open-ended dialogues enabling free small talk (see Figure 4.1).

DB-CALL is situated within the interactionist perspective on SLA, relying on Long's Interaction Hypothesis (1996), where dialogue with the CA provides the student with spoken input, the opportunity to express output, and instances of negotiation of meaning and feedback (see Section 6.1.1). In line with a system's interactional design and level of constraints on meaning and form, DB-CALL may allow spontaneous speech practice (fluency) and the automatization of knowledge when interacting in the target language to engage in the meaningful production of utterances (DeKeyser, 2007). It enables the idea of dynamic cooperation of implicit knowledge through the input of L2, together with some explicit knowledge (e.g., grammar) to assist L2 learning and development that is gradual over time (the weak-interface position, see Section 4.2; Ellis et al., 2009). Similarly, the four interrelated strands of L2 learning suggested by Nation (2007), consisting of a focus on meaningful input, meaningful output, explicit language instruction, and fluency development, collectively contribute to a comprehensive and balanced approach to L2 learning encompassing aspects such as production, comprehension, and fluency.

From a pedagogical perspective, DB-CALL applies TBLT, which entails students actively using the L2 to complete a task (Ellis et al., 2020; González-Lloret, 2017). A task is an organisational unit defined as an activity during which the student is supposed to employ L2 communication to solve problems and achieve goals linked to a real-world situation as a foundation for L2 learning. The students must reach a set goal, focusing on meaning, in contrast to the exercise approach, which focuses on form or vocabulary (VanPatten & Benati, 2015, p. 195). According to Ellis et al. (2020), a task is:

a piece of classroom work that involves learners in comprehending, manipulating, producing, or interacting in the target language while their attention is focused on mobilising their grammatical knowledge in order to express meaning, in which the intention is to convey meaning rather than to manipulate form. (p. 4)

By engaging students in meaningful speaking activities and problem-solving tasks, TBLT promotes the development of L2 speaking skills within an authentic context.

In this thesis, the term “task” was interpreted accordingly, as an activity in which the students used their existing L2 knowledge to communicate and fulfil a specific task. It is used interchangeably with the terms “goal”, or “solving a problem”, such as ordering food at a restaurant or buying a train ticket. The focus is on the spoken modality practised in two task-oriented SDSs with humanlike ECAs, which are further described and compared based on the presented conceptual framework of DB-CALL (see Section 7.1.4). Next, there follows a presentation of the central ideas and principles of SLA and CALL which frame this thesis and selected SDSs.

## 4.2 SECOND-LANGUAGE ACQUISITION AND COMPUTER-ASSISTED LANGUAGE LEARNING

The field of SLA is interdisciplinary, drawing upon, among others, socio-cultural, linguistic, psycholinguistic, cognitive psychology, and neurolinguistic perspectives on the competencies and skills involved in L2 or L3 learning and development. Issues of interest are the four language skills (reading, writing, speaking, and listening), individual differences, assessment, and the setting (Gass & Mackey, 2012). One topic of argument is the role played by explicit knowledge in L2 use during the learning and development of communicative ability (Interface Issue), with three claimed positions: (i) non-interface, (ii) strong-interface, and (iii) weak-interface (Ellis et al., 2009). The non-interface position claims that, in order to develop implicit knowledge, the teaching must be implicit (Krashen, 1985). The strong-interface position stresses that L2 learning is equivalent to any other learning and that anything you also learn explicitly (e.g., a grammar exercise) is transformed through practice into implicit knowledge (automatisation). Pedagogically, this is often summarised into three steps as pedagogical strategy: (i) presentation, (ii) practice, and (iii) production (DeKeyser, 2007). Finally, the weak-interface position, a dynamic mixing of the previous ideas, stresses that students must have implicit knowledge through possible input for L2 development, together



with some explicit knowledge to assist L2 learning and development that is gradual over time (Ellis et al., 2009).

Adding these various SLA interests into the CALL field has occupied researchers since the 1960s, along with evolving technologies (Heift & Chapelle, 2012; Godwin-Jones, 2021). Consequently, SLA findings have also had an impact in the field of CALL (Heift & Chapelle, 2012; Levy, 1990, 1997), explained in the various approaches taken in CALL (Bax, 2003, 2011; Warschauer & Kern, 2000). According to Warschauer and Kern (2000) and Warschauer (2004), CALL research has developed over the years in phases, alongside three language-learning paradigms: (i) structural, (ii) communicative, and (iii) integrative CALL.

- (i) The first phase, known as Structural CALL (Warschauer & Kern, 2000), was developed during the 1950s–1980 based on a view of language as a structural system (e.g., grammar, phonology) and behaviourist theories of learning that emphasise imitation and repetition (Skinner, 1968). This resulted in grammar-translation and audio-lingual methods, emphasising mechanical and pattern-based practice (Richards & Rodgers, 2001). Here, the role of technology is more of a tutor, employing repetitive drill-and-practice programmes or adaptive tutoring systems (Walker & White, 2017). Structural CALL is still often applied as an approach in digital tools because it has the potential to develop L2 skills, such as writing, with the principal objective being accuracy and systematic language instruction with positive reinforcement.
- (ii) The second phase, commonly known as Communicative CALL, developed during the 1980s–1990s, drawing upon the underlying view of language learning as constructed in the student’s mind. It uses the methodology of communicative language teaching (Littlewood, 1984). This phase focuses on the development of communicative competence, as defined by Hymes (1972), which encompasses both knowledge of grammar and vocabulary and the ability to use language effectively in real-life situations, taking into account social and cultural factors. Hence, there is a focus on the practical application of language and lan-

guage learning through interaction/communication with others. Computers are primarily used for communicative exercises, with a focus on developing fluency. This phase was also inspired by a constructivist view of learning, stressing the need for the student to engage in communicative exercises and experience in order to learn (Papert, 1993), where the target language can be used as a tool in a meaningful context to solve problems. The use of technology emphasises the need for student-centred activities. It also includes intelligent CALL (ICALL), which focuses on feedback and AI, providing opportunities for the construction of individual knowledge in meaningful audio-visual contexts and specific domains, such as grammar, with individual support in so-called intelligent tutoring systems (Li, 2017). Some systems also provide adaptive sequencing and the individualisation of instruction.

- (iii) The third phase, known as Integrative CALL, developed during the 21st century in the teaching paradigm of content-based instruction (Kern & Warschauer, 2000). The aim is to integrate students into authentic environments for interaction in the target language through multimedia and the internet. This phase includes activities to interact via computer in so-called computer-mediated communication (Blake, 2013) focusing on meaning before form. When communication is seen as the goal of learning, students must have the opportunity to engage in everyday-life interactions. The principal objective is agency, for students to enjoy more autonomy and be agents in their own learning process (Warschauer, 2004). This phase includes CALL practice aimed at developing language skills that employ technology as a mediational tool regardless of time and place (Li, 2017). It is based on a sociocultural perspective on learning (Vygotsky, 1978), developing language through the “negotiation of meaning in dialogues with the target language in the real world” (Li, 2017, p. 20).

This way of grouping CALL's development into three phases has been criticised for being portrayed as linear by Bax (2003), who instead outlined a nuanced and more dynamic model that takes into account the various elements that have shaped the evolution of CALL, such as the physical location of computers, the role of teachers, the type of activity, and feedback. The physical location was considered relevant because early CALL systems were designed to be used in limited computer rooms or labs, or in classrooms. This contrasts with more recent developments in CALL, which have emphasised the importance of mobile and ubiquitous computing, enabling students to access L2 digital tools from anywhere, at any time. Bax's (2003) first approach was (i) Restricted CALL, referring to restrictions on tasks, questions, feedback, and content that was part of the programmed system. His second approach was (ii) Open CALL, referring to open-ended interactions with computers (as in Communicative CALL). Bax's (2003) final approach was (iii) Integrated CALL, signalling a phase in which technology has become part of everyday life and teaching, and is fully normalised (Bax, 2011). This last approach shares the sociocultural view of learning described in Warschauer and Kern's (2000) last phase, Integrative CALL.

Today, the field of CALL can be seen to involve "any environment in which a learner, alone or collaboratively with peers, uses technology in a second or other language" (Heift & Chapelle, 2012, p. 556), still having the main aim of "studying how technology can help learners develop their language ability" (*ibid.*, p. 557). With the rapid development of technology, the research is moving towards also including other devices, such as smartphones and tablets, in technology-enhanced language learning (TELL) "as part of the environment in which language exists and is used" (Walker & White, 2017, p. 9). The focus is on communication and interaction in collaboration in a social constructivist sense of using language as a medium in knowledge construction (Blake, 2013). However, despite the extensive integration of technology into daily life, educational technology has not yet reached complete integration into L2 teaching (*ibid.*), or achieved full normalisation (Bax, 2011). The different ways of viewing L2 learning and the theoretical underpinnings of L2 learning and technology may affect the focus of CALL studies, and could also explain the design of educational digital tools.

In this thesis, the selected SDSs Enskill (Alelo, 2022) and ImmerseMe (n.d.) are mostly associated with integrative CALL, although earlier phases and approaches are still valid in various combinations and types of setting in educational technology. As within the phase of integrative CALL, this thesis refers to the employed technology as digital tools in general, and conversational AI, referring to SDSs with ECAs, specifically. The employed SDSs are interpreted as a mediational tool for student-to-CA interaction in the target language and as a tutor or instructor of the student. The educational environment for Studies 2–4 consisted of two types of virtual learning environments that can be displayed on the screen of any digital device for practising L2 speaking, activating various complex speaking skills in interplay with emotional factors, as presented in the next section.

### 4.3 INTERPLAY OF COGNITIVE AND AFFECTIVE FACTORS

The research literature shows that, when learning and developing the skills to speak a language, various cognitive and affective factors interact simultaneously, where affective refers to a broad range of feelings, emotions, attitudes, and moods that influence a student's behaviour and experiences (Goh & Burns, 2012). Swain (2013) has highlighted the importance of considering emotions as an affective response in L2 learning, something that has recently been addressed in SLA research and is highly relevant for the learning and development of speaking skills (Li, 2017). Students are more motivated to engage in learning activities when they experience positive emotions such as happiness, enjoyment, excitement, satisfaction, etc., in contrast to negative emotions such as anger, boredom, or anxiety, which often lead to disengagement (Sang & Hiver, 2021). Goh and Burns (2012) have stressed that emotional language anxiety due, for instance, to social pressure, a threatening environment, or a lack of motivation (e.g. lack of confidence or no ideas about the future use of an L2) may negatively affect students' willingness to communicate (see MacIntyre et al., 1998), and hence their engagement in the speaking activity. Emotions can also be related to the extent to which it feels meaningful to practice speaking in this way, which is closely linked to a pragmatic approach to learning,

emphasising first-hand meaningful experiences for learning and problem-solving theory. L2 learning is hence closely linked to doing and practising hands-on activities related to the goal of the skill (Dewey, 1959); in this thesis, speaking skills. The interaction provided in SDSs is in line with TBLT (Ellis et al., 2020), featuring situations around a practical problem or a task for the student to solve together with the CA in the simulated everyday-life dialogue, for example, ordering at a restaurant, which is supposed to relate to the student's interests and life (Nation, 2007).

In this thesis, students' self-reported emotional experiences were interpreted as automatic reactions to what happened inside them while they were practising L2 speaking with an ECA in an SDS, aspects that trigger particular feelings and might influence language learning (see Chapter 7). Regardless of the underpinning theories, some commonly established key principles (Ellis, 2005; Li, 2017; Nation, 2007) are considered important for effective L2 learning and sustaining students' engagement, as presented next.

#### 4.4 PRINCIPLES FOR EFFECTIVE L2 LEARNING

Acknowledging the complexity of L2 learning, Li (2017) has suggested six theoretically and empirically based key principles for effective L2 learning, which are also beneficial within DB-CALL. These comprise:

- (i) authentic input
- (ii) conscious noticing on form
- (iii) opportunity for interaction
- (iv) in-time and individualised feedback
- (v) low affective filter
- (vi) an environment where language can be used

These principles must not be seen as an exhaustive list of principles that may influence learning. In this thesis, they have informed the content of the instruments. Furthermore, these principles reveal the importance of using the target language actively (Ellis, 2005). Additionally, it is also beneficial to participate in social interaction, preferably with a more knowl-

edgeable interlocutor, an expert (Lantolf et al., 2014; Vygotsky, 1978), who provides adequate scaffolding possibilities, enabling the student to be prepared for everyday-life situations as prompted in TBLT (Ellis et al., 2020; Long, 2015). Concerning the CALL interaction, students “are expected to input, feedback and interact in a way similar to what they would experience in face-to-face interaction” (Li, 2017, p. 53). Next follows a brief presentation of these principles, here clustered into three subsections, labelled: (i) authentic input and focus on meaning, (ii) noticing on form and feedback, and (iii) low affective filter. Interaction and the Interaction Hypothesis are further discussed in Chapter 6.

#### 4.4.1 AUTHENTIC INPUT AND FOCUS ON MEANING IN INTERACTION

There is consensus in the research that, in order to learn a language, students must actively use it in an authentic communicative context and engage in interactive activities that focus on meaning (Ellis et al., 2020; Li, 2017; Mitchell et al., 2012). In SLA, authentic is used in the sense of material, tasks, or contexts that closely resemble real-world, natural language use (Li, 2017; Long, 1996). It emphasises the use of genuine, unmodified language and real-life communicative situations that students may encounter in the target language outside the language classroom. The authenticity of learning materials, experiences, and environments has been identified as beneficial for students’ engagement (Egbert et al., 2009), and the importance of authentic and compelling input for language learning has also been highlighted (Li, 2017; Long, 1996). The dialogues must be around authentic tasks to solve while maintaining a fluent dialogue in the target language. Hence as a condition for practising L2 speaking with authentic input, there is hence a need for an environment in which students have an opportunity to interact, including input, output, feedback, and the negotiation of meaning in everyday-life scenarios that are relevant to the student (Li, 2017).

In this thesis, the phrase “everyday life” is used in parallel with “authentic” because the classroom context can also be regarded as an authentic part of students’ “real life”. The speaking activities simulated in the SDSs aim to be identical to everyday-life scenarios outside an educational con-

text, including social aspects and natural language use. The virtual environment of the SDS, animated or filmed, including the ECAs, can engage and immerse students to various extents, and give them a sense of physical presence, of really being there, to make it feel authentic.

#### 4.4.2 NOTICING OF FORM AND FEEDBACK

Although the main focus in L2 learning in interaction is on meaning, it has been suggested that there must also be some conscious focus on form in its natural occurrence, with students being provided with opportunities to notice linguistic elements (accuracy) such as lexical, grammatical, and phonological aspects (Long, 1996). When students focus on form, as also recognised in the negotiation of meaning in L2 learning, this is considered beneficial for learning and development. The role of feedback is a widely researched topic within SLA and can be realised through either implicit or explicit individual information provided to the student (Loewen, 2012), such as the ill-formedness of their L2 production, to guide improvement in their speaking skills (Li, 2017).

In line with Long (1996), implicit feedback can be provided in the form of recasts or reformulations in the interaction, which is given during the spontaneous communication while the students and interlocutor modify their utterances to improve understanding. As an example of explicit feedback, there is corrective feedback given in the form of comments on mistakes, either with the correct form already given or prompting students to correct it themselves (VanPatten & Benati, 2015). Additionally, there is summative feedback, which focuses on the evaluation and assessment of the student's performance against a predetermined set of standards or criteria (Goh & Burns, 2012)

Within L2 education, it can be a challenge for teachers to consistently provide sufficient feedback within the given time for a class of students (Li, 2017). In terms of speaking, which is transient, giving feedback requires listening to each speaking student in real-time, or recording the students' speech. Therefore, CALL could be a supportive source for providing students with "timely, accurate and individualised feedback" (Li, 2017, p. 159).

In this thesis, there is a combination of instant feedback, provided by the ECA during the spoken interaction in the turn-taking, or eventual breakdowns in the communication and transcripts displayed in real-time alongside the spoken utterances. There is no explicit feedback on pronunciation, but summative feedback is provided in written information about how well the learning objectives were fulfilled. Formative feedback is provided as individually adapted practical grammar exercises after the speaking practice in one of the SDSs, to aid in the improvement of the skill and to strengthen the student's confidence in speaking the target language.

#### 4.4.3 LOW AFFECTIVE FILTER

Personal internal variables, such as the level of anxiety, confidence, and motivation about speaking in the target language affect students' involvement in speaking activities and their willingness to speak (Goh & Burns, 2012; Li, 2017; MacIntyre et al., 1998; Mitchell et al., 2012). These key variables are sometimes referred to with the metaphor of the "affective filter" (Krashen, 1985; Li, 2017, p. 35).

Anxiety can be provoked by obstacles such as fear of saying something wrong, not understanding, or not being understood due to a lack of vocabulary or a non-native accent in the target language. Students might, for instance, feel too exposed to other students to speak and/or afraid of being judged (Walker & White, 2013). Suggested methods for reducing affective filters have been incorporated by TBLT in the form of appropriate language use and clear learning objectives (Goh & Burns, 2012) and technological assistance (Li, 2017). Authentic communication tasks in so-called "whole-task practice" (Goh & Burns, 2012, p. 136) are provided to increase the communicative need and the personal motivation to keep the conversation going.

Social pressure in the classroom might also negatively affect speaking ability, so a safe environment without threat is another important factor for developing speaking skills. Furthermore, students could be helped by being guided progressively with feedback and support through scaffolding, being encouraged to reflect upon the development of their speaking skills, and self-regulated learning through monitoring and evaluations. There are cognitive phases; for instance, conceptual preparation with themes so



students can produce spontaneous speech within familiar situations and topics (Goh & Burns, 2012). This is supported by content knowledge that could be prepared beforehand and includes rehearsals and task repetition to build students' confidence.

Motivation is another trait exhibiting individual differences; this is a complex construct representing a whole research field (Ushioda & Dörnyei, 2012). Motivation is one variable among many interrelated variables that have implications for L2 learning, and displays positive correlations with willingness to communicate (MacIntyre et al., 1998). Hence, teachers must strive to reduce anxiety within the learning situation and its materials. Additionally, motivation can also be helped by autonomous and self-regulated students taking control over their learning in an active way in an environment that is less restricted and threatening than traditional classrooms (Li, 2017).



## CHAPTER 5

# RELATED RESEARCH

The purpose of this chapter is to outline the present state of research in relation to conversational AI applied to institutional L2 education from teachers' and students' perspectives in order to frame how this thesis can contribute to the field with new knowledge to the field. This chapter includes four overarching sections. The first presents a synthesis of the central studies about L2 teachers and digital tools, including conversational AI, as experienced from teachers' perspectives, finally zooming in on Swedish teachers. This section also includes work conducted with teachers in relation to the SDSs employed in this thesis, Enskill and ImmerseMe. The second section presents the central studies in which students have practised L2 speaking skills in SDSs. The third presents studies that are central to speaking experiences with CAs. Since there is some overlap with research about pedagogical agents, virtual humans, and ECAs in terms of the social aspects of generating social agency, such studies are also included. Finally, in the fourth section, central aspects of the previous research are briefly reviewed.

### 5.1 L2 TEACHERS' VIEWS AND EXPERIENCES

The potential of emerging technologies such as AI and educational robots in education is widely acknowledged, but simultaneously there are critical reflections and concerns around the future role of teachers, ethics, economic interests, and the need for more knowledge about AIED (Hrastinski et al., 2019; Humble & Mozelius, 2022). When Haelermans (2017) explored why digital tools were not optimally used in education, she found that teachers erected barriers to technical change due to internal beliefs,

insufficient “know-how”, or lack of time to put them into practice or training. She stressed the importance of teachers if digital tools are to be useful in education, for which her classroom studies in Dutch secondary schools indicated significant potential, especially in maths and languages and for low-proficiency students. A comparison between traditional teaching and teaching enhanced with interactive digital tools revealed the usefulness of the latter; for example, in practising skills (e.g. language spelling and grammar), effectively providing feedback, and individualisation.

According to Li (2017), L2 teaching and methodology must keep up with the needs of today’s new generation of students and their daily interactions with technology, and take advantage of the rapidly emerging technology supporting the learning of various L2 skills. The reasons mentioned include: to improve students’ engagement, enhance language learning and facilitate understanding, provide a context for using the target language, and also, as an effect, to reduce teachers’ workload and save time in the classroom. The actual integration and use of technology are said to be influenced by various factors, such as teachers’ attitudes, views, and pedagogical beliefs; systems of policy, practice, and curriculum; resource accessibility; and digital competence and confidence, as shown in various smaller studies. For instance, Çelik and Aytin (2014) found that Turkish elementary and high-school teachers ( $N = 6$ ) reported positive views, sufficient digital competence, and willingness to apply digital tools when teaching English. However, practical obstacles to using them, such as lack of computers and unstable Internet connections, were given as the main factors inhibiting their actual use in the L2 language classroom. The implementation process has been highlighted as challenging to apply in a way that meets the curricula and students’ individual needs and likes (Soto et al., 2020).

Timpe-Laughlin et al. (2020) have explored the experiences of 16 teachers in an intensive English programme at a US university who reported their user experience of an SDS. The teachers’ ratings, revolving around their level of engagement and the perceived usefulness of SDS-based, goal-oriented speaking tasks, revealed positive views regarding the potential for SDS to be used for speaking practice in L2 classes, especially when applied in a flipped classroom and for diagnostic purposes.

In another usability study, van Doremalen et al. (2016) explored how teachers ( $N = 9$ ) combined with 15 students used the ASR-based SDS

prototype DISCO, providing feedback on speaking performance. The results indicated positive views about its potential for enhancing speaking through skill-specific practice, with ECAs providing appropriate feedback, but some concerns were raised about the difficult vocabulary used. They thought that, since students were stimulated to speak more, they subsequently became more comfortable and that spoken interaction added value compared to written interaction. Important conclusions were that this kind of system must add different training strategies to address the errors made by an individual student. Additionally, the study showed that content and authenticity with relevance to students are crucial factors for the application of the system. One conclusion is that there is “no one-size-fits-all CALL system” (van Doremalen et al., 2016, p. 845) and that it must be possible to modify the content.

Belda-Medina and Calvo-Ferrer (2022), applying a mixed-methods approach, investigated the level of satisfaction and perceptions regarding the integration of conversational AI into language education among Spanish and Polish trainee teachers ( $N = 176$ ) who interacted with ECAs for a month. Their findings showed that gender had no effect on the participants’ satisfaction in relation to the linguistic level of the ECAs. In contrast, the qualitative findings did reveal gender effects on satisfaction, with female participants being more attentive to design and content. Overall, the paper reported generally positive findings regarding the integration in terms of ease of use and attitudes towards integration, although with only moderate knowledge about conversational AI and interest in using it in future education. In sum, previous short intervention studies have often relied on interview data or surveys with mostly small samples, providing a state of knowledge with promising results internationally.

In Swedish L2 education, in relation to speaking skills a study about the actual pedagogical use of digital tools by L2 trainee teachers ( $N=12$ ) has shown that a lack of pedagogical models from their own educational background led to a sense of low self-efficacy, knowledge, and belief in their capacity to act as a teacher. This has implications such as an expressed need for knowledge and reflected experiences (Nilsson, 2017). It is also worth mentioning a comprehensive Swedish research project conducted during 2016–2019, which examined the interactions and alignment between teaching, assessment, and learning (TAL) in relation to modern-language

speaking proficiency (TAL, n.d.). It investigated framing factors at various levels, such as society, school, attitudes, motivation, and background among lower-secondary students in the Swedish school context (Erickson et al., 2022). A questionnaire was completed by school principals ( $N = 141$ ) and teachers ( $N = 315$ ), complemented by conversations with teachers and students and recordings of the spoken performance of students ( $N = 150$ ). Although teachers' perspectives on language education in compulsory school were the focus of the study (in this era of digitalisation), there were no specific questions about teachers' views or experiences of using digital tools in relation to speaking skills. In sum, there is a need for further studies investigating Swedish language teachers' views and experiences of using digital tools in order to practise speaking.

#### 5.1.1 STUDIES IN RELATION TO THE EMPLOYED SDSs

L2 teachers' perspectives on using Enskill (used in Studies 2 and 4 of this thesis) in their teaching, is known through spin-off results from the few existing student-centred studies. In Johnson's (2020) case study of Mexican university students ( $N = 107$ ) using the SDS Enskill regularly for two months, as an integrated part of their curriculum, the involved teachers reported increased fluency of speech and confidence in their students after practising speaking in this system. It generated more communicative classes, where the students engaged in conversations. As perceived by the teachers, one possible explanation for these benefits was that, while practising English in the system, the students received personalised instruction and therefore came to class better prepared.

Focusing on teachers' perspectives on the benefits and drawbacks of VR, and the challenges and effects of practising speaking in the SDS ImmerseMe (used in Study 3 of this thesis), Tran (2021) empirically examined the opinions of L2 teachers ( $N = 105$ ) in an international online survey, complemented with focus group interviews. The clearest benefits for developing speaking and listening skills were gaining access to authentic culture, real-life conversations, and creating a meaningful context resulting in the students gaining confidence. The main drawbacks were in terms of VR being time-consuming due to having to redesign lessons, and its limitations in replacing classroom interaction. The main challenges in the

implementation of VR were a lack of the required facilities (hardware and software), a lack of digital skills training, and the insensitive functionality of ASR, which provoked frustration among both teachers and students. The conclusions of the study were that, generally, teachers needed to improve their digital competence to handle technical issues, have an “adaptive mindset” with backup plans in case of technical issues, and collaborate with colleagues when redesigning lesson plans, including VR, to be optimally beneficial for the students (Tran, 2021, p. 64).

## 5.2 STUDENTS PRACTISING L2 SPEAKING SKILLS IN AN SDS

The implementation of conversational AI in education is still in its infancy (Zhai & Wibowo, 2023), although previous research has shown that practising L2 speaking in an SDS benefits students’ speaking skills (e.g., Bajorek, 2018; Golonka et al., 2014; Utami et al., 2021). While most studies focus on university students and quantitative data, a few examine younger students’ self-reported experiences. This section presents the relevant studies, organised by findings and connected to SLA principles (e.g., authentic input and feedback). It highlights the strengths and weaknesses of SDSs based on previous research, thus setting the stage for this thesis.

### 5.2.1 BENEFICIAL FOR L2 SPEAKING, LISTENING, AND VOCABULARY

Practising pronunciation using ASR has been found to facilitate pronunciation improvement (e.g., Bajorek, 2018; Kukulska-Hulme & Lee, 2020). SDSs can provide students with meaningful and authentic language practice (Berti, 2020), with the opportunity to interact with CAs and receive immediate feedback from the system on their performance from the system (Bibauw, 2022; van Doremalen et al., 2016; Hassani et al., 2016). Studies using SDSs to enable users to engage in everyday-life situations have been shown to facilitate vocabulary acquisition (Divekar et al., 2021; Kaplan-Rakowski & Wojdyński, 2018) as well as the development of speaking and listening skills (Bajorek, 2018; Blake, 2017; Johnson, 2019a; Li, 2017). Hsu et al. (2021) conducted an experimental design study over

four months and showed that Taiwanese university students ( $N = 100$ ) who conversed with a task-oriented chatbot system reduced their fear of speaking English and improved their spoken ability. These results are in line with the findings of Kim et al. (2021), who, based on pre- and post-tests examined how Korean university students' ( $N = 49$ ) speaking skills were positively affected by speaking with the AI chatbots Replika, Andy, and Google Assistant, in particular their pronunciation, intonation, stress, and fluency. In their systematic overview of SDSs, Bibauw et al. (2019) showed that most systems demonstrated enhanced speaking skills, especially accuracy. Timpe-Laughlin et al. (2022) examined the utility of an SDS (HALEF) using semi-scripted role-plays with feedback for L2 learning based on data produced in a post-task user perception survey completed by a group of students ( $N = 47$ ) of English at the tertiary level in the USA. The system enabled conversation through questions and answers and the negotiation of meaning. There were supportive results for SDS tasks as an alternative for effectively practising L2 speaking and producing data about L2 speaking.

Empirical studies have shown that SDSs provide opportunities for practising L2 speaking in simulated everyday-life scenarios through human-to-machine interaction where maintaining a dialogue is also the task (Bibauw et al., 2022; Timpe-Laughlin et al., 2020). The students were able to actively use the target language in a contextualised way, by repeatedly engaging in face-to-face interaction with CAs with flexibility in time and space (Cucchiarini & Strik, 2019; Huang et al., 2021), without travelling abroad. It has been found that SDSs promote self-regulated learning because students can control the pace, content, and support of their language practice (e.g., Kaplan-Rakowski & Wojdyński, 2018; Meri-Yilan, 2019) and take the initiative in constructing their learning (Chong & Reinders, 2020). Studies have shown support for blended learning settings where face-to-face methods are integrated and mixed with DB-CALL systems in language education (Grgurovic, 2017).



### 5.2.2 LOW-ANXIETY ENVIRONMENT, GAINS IN CONFIDENCE AND WILLINGNESS TO COMMUNICATE

Practising speaking in a low-anxiety environment (Timpe-Laughlin et al., 2020) has shown positive results, including reductions in anxiety, fear, and embarrassment when speaking, with increased confidence and willingness to communicate in the target language (Ayedoun et al., 2015; Divekar et al., 2021; Jeon, 2022; Papin, 2018; Wang et al., 2017). In an earlier study, Johnson and Wang (2008) reported that users' self-efficacy (confidence in their ability to learn, speak, and understand the target language) increased when using a system that provided appropriate feedback.

SDSs provide more enjoyable and less stressful speaking practice than speaking with a human interlocutor in a classroom environment (Bashori et al., 2020). Bashori et al. (2021) investigated secondary school students ( $N = 232$ ) in Indonesia in a quasi-experimental study based on questionnaires, vocabulary tests, and some complementary interviews, and the results also indicated lower levels of anxiety among the students when practising speaking in an SDS. Grant et al. (2013) showed how students could improve their listening and speaking skills in simulations representing a safe VR environment to prevent them feeling anxious. Practising conversation with chatbots in an enjoyable and non-threatening environment is also suggested in systems like Mondly and Duolingo, in Eggbum (text-bound), with a grammar-bot in Memrise (Alm & Nkomo, 2020), and with pre-filmed native speakers in the 360-degree environment in ImmerseMe (Berti, 2020).

### 5.2.3 EFFECTIVENESS VERSUS BIAS

Researchers have addressed how effective SDSs are for L2 learning, and Bibauw et al. (2022) have found that corrective feedback and session spacing were moderators for effect. Based on quantitative measurements, their results indicated a medium effect on L2 proficiency development and established “significant effects for lower proficiency learners, and on vocabulary, morphosyntax, holistic proficiency and accuracy” (p. 1). The tendency of higher effects for low-proficiency students was also found in a study by Huang et al. (2008). In terms of effectiveness, empirical stud-

ies focusing on attitude and cognitive aspects have shown a greater positive impact on beginners and lower-intermediate language learners (Kim, 2016). In contrast, Sydorenko et al. (2018) suggested that advanced-level students benefit more from such interactions due to their ability to listen and understand the CA's utterances. Additionally, it has also been observed that low-proficiency students in English were less recognised in human-to-machine interactions with other types of conversational AI, such as intelligent personal assistants (Moussalli & Cardoso, 2020). Fryer et al. (2019), in an extension study, suggested that, based on regression and content analysis, the surveyed students' learning experience with chatbots for L2 learning was predicted by prior interest in human conversation partners, prior language competence, and task interest.

Criticism directed at DB-CALL research focuses on the reliance on study designs that carry a risk of bias regarding a novelty effect when technology is first launched to students. Students' short-term initial engagement and performance may be positive, only to then decrease as they become increasingly familiar with the SDS (Huang et al., 2021; Fryer et al., 2019). Empirical studies with pedagogical evaluations in reviews, observational, survey, and effectiveness studies have been of varied scope and quality and have mostly not achieved statistical significance due to small samples and the short time spent on the task. One possible source of bias is that developers and companies are sometimes involved in the research studies, and few of the systems researched have made it beyond prototyping.

One early system that has actually been used publicly is SPELL (spoken electronic language learning), which is described as a user-tested CALL system for classroom use (Anderson et al., 2008). It offered conversational scenarios from everyday life by implementing virtual worlds, humanlike ECAs, and ASR. After investigating and evaluating SPELL in several phases, including with various nationalities and target languages (Italian, Japanese, English, and French), the results suggested that the acoustic models interpreted by the speech recognition system had been trained with native speakers but needed to recognise the spoken language of non-native speakers, which consequently led to many breakdowns in the conversation. Based on usability questionnaires, verbal interviews, and measured ASR accuracy, a motivational boost was noticed among high-

school students, who felt satisfied with their performance in the system, and mostly reported it as useful and effective for L2 learning (Anderson et al., 2008).

#### 5.2.4 ATTITUDES, ENGAGEMENT, AND MOTIVATION FOR L2 LEARNING

Studies regarding students' attitudes towards and evaluations of SDSs have shown overall positive but varied results for the examined systems in the systematic review by Bibauw et al. (2019). Kaplan-Rakowski and Wojdyński (2018) explored in their laboratory study the experiences of 22 adult male Polish students interacting in English in everyday-life scenarios with ECAs in the SDS Mondly VR. The results showed that most respondents had positive attitudes about the impact of VR-based assisted language learning and reported high engagement with the speaking activity due to feeling immersed with little distraction.

Jeon (2022) explored the impact of chatbots on students' motivation for L2 learning. The study analysed the qualitative data produced in system-generated interaction logs and post-trial interviews, focusing on how Korean primary-school students ( $N = 36$ ) engaged with chatbots in goal-oriented conversations for a duration of 16 weeks. The study explored the inherent possibilities of language learning with the chatbots, considering various aspects, such as pedagogical, technical, and social factors. The findings indicated that these factors influenced learning outcomes differently, acting either as limitations or opportunities for learning. Technology-related factors have also been examined and found to affect students' experiences within SDSs, sometimes causing frustration and having a significant impact upon students' engagement (e.g., Alm & Nkomo, 2020; Johnson, 2019a). An SDS's constraints, such as problems in understanding several utterances in a row, have been identified as a challenge, together with a restricted conversational path due to the lack of negotiation of meaning that takes place during human-to-human interactions (Bibauw et al., 2019).

### 5.2.5 GOAL-ORIENTED SYSTEMS, FEEDBACK, AND SCAFFOLDING

In their systematic review, Bibauw et al. (2019) highlighted positive evidence for “goal-oriented systems”, which are sometimes also referred to as “task-oriented” prompting of “contextualised dialogues” (p. 31), where students express themselves within the given context, supported by the system’s feedback and scaffolding. Corrective feedback which is permitted in highly constrained SDSs seems to enable higher L2 learning gains than systems that allow more spontaneous speech (Bibauw et al., 2022). Tai et al. (2022) conducted an experimental study with a control group in which they interviewed adolescent Taiwanese students who interacted with intelligent personal assistants in English. The researchers focused on the impact of the feedback presentation mode on learning speaking proficiency and found that the on-screen text, graphics, and narration were crucial.

Ayedoun et al. (2020) explored Japanese university students’ ( $N = 60$ ) willingness to communicate in terms of confidence, anxiety, and desire to communicate in the target language (English) with an ECA in various scenarios by implementing scaffolding to support and motivate them to engage in the interaction, so-called motivational scaffolds. The focus was on scaffolding in terms of conversational strategies through affective backchannels and task-based communicative strategies and how to fade them out in relation to the students’ preferences and needs. The results suggested that progressive and fading motivational scaffolds in the system were dependent on the level of the students’ willingness to communicate, and addressed the need to develop adaptive systems to meet various students’ expectations.

### 5.2.6 VR IMMERSION IN SDSs

With the addition of VR, the student can become further immersed in the speaking experience, with increased focus due to the sense of physically being there while removing geographical limitations, and having a feeling of presence and a sense of authenticity (Kaplan-Rakowski & Wojdyński, 2018; Wang et al., 2017). VR is thus “supporting authentic learning in the 21<sup>st</sup>-century language classroom” (Utami et al., 2021, p.133), whereof

a synthesised overview shows five benefits for learning, consisting of: student motivation, writing skills, oral production by providing meaningful new activities, listening skills, and academic achievement (ibid., p. 133).

For instance, Divekar et al. (2021) explored an SDS with extended VR and CAs, which showed statistically significant improvements in students' vocabulary learning, listening comprehension, and speaking ability. The immersion, involving senses such as sight and hearing, makes the student feel present and encourages positive feelings, enjoyment, and the development of speaking ability (Kassim et al., 2019). VR can also bring the culture of the target language closer to the student (Berti, 2019; Johnson & Valente, 2009; Utami et al., 2021). Drawbacks identified with high-immersion VR include eye tiredness and headaches (e.g. cybersickness; Dhimolea et al., 2022; Kaplan-Rakowski & Wojdyński, 2018).

Two variants of VR are applied in the SDSs selected for this thesis. In Enskill, the virtual environment is animated using virtual humans as ECAs, and ImmerseMe offers spherical 360-degree videos with pre-filmed native speakers as ECAs. The next subsection outlines some of the findings from studies of importance to this thesis conducted with these two SDSs, Johnson (2019a) which is relevant to Studies 2 and 4, and Soto et al. (2020), which is relevant to Study 3, where the latter was published after the current study was completed.

### 5.2.7 STUDIES IN RELATION TO THE EMPLOYED SDSs

When it comes to practising L2 speaking skills in Enskill (see system description in Section 7.1.4.1), by enabling interactions with ECAs, as employed in Studies 2 and 4, a few previously published studies have been conducted (Johnson, 2019a, 2021, and some presented in white papers, e.g., Johnson, 2019b, 2020), all in close conjunction to the company behind the system. Overall, the results have shown improved speaking skills, increased confidence, and greater fluency.

Using snapshot evaluations, Johnson (2019a) reported on two case studies with university students ( $N = 111$ ). One group of B-level (intermediate) students studying English in Serbia used the system for three weeks as homework assignments, and the second group was in Croatia. The students' self-reported attitudes and experiences of practising in the

system were collected in a questionnaire using a five-point Likert scale and open items for evaluating the system. System-generated metrics in terms of raw understanding rate complemented the self-reports in order to analyse the students' performance. The focus was also on identifying linguistic errors (e.g., words, grammar, and pronunciation). The students reported that it was a good, easy, and fun system for practising speaking English, and most said that the interactive conversations were "amazing" (p. 438). However, there were mixed opinions in terms of whether the interactive conversations helped to develop their English speaking and listening skills. Some participating students reflected upon the limited choices of dialogue subject and bugs reported in the system. Long-term studies for measuring effectiveness were suggested to see if the developed skills were retained over time. Enskill has recently also been used in a case study with community health workers ( $N = 57$ ), who rapidly improved their communication skills after practising in simulated scenarios with CAs acting as simulated patients (Johnson, 2021).

When it comes to practising speaking skills in ImmerseMe, with pre-filmed native speakers, the system used in Study 3 (see system description in section 7.1.4.2), some previous studies have been conducted, mostly highlighting the benefits of immersive VR for enabling meaningful contextualised L2 practice (Papin, 2018; Soto et al., 2020). Central studies for this thesis include a review (Berti, 2020) about this low-immersive (Kaplan & Gruber, 2019) version of interactive, VR-based learning relying on ASR for learning useful phrases. It shows valuable access to native speakers of the target language outside the classroom context, and empirical research indicates that L2 learning can be promoted through 360-degree videos (Meri-Yilan, 2019). Higher education learners' experiences of interaction and learning depending on immersion confirmed the earlier theoretical model of Ai-Lim Lee et al. (2010), and were measured in terms of factors such as "usability, presence, motivation, cognitive benefits, control, active learning, reflective thinking and learning outcomes" (Meri-Yilan, 2019, p. 275). It was found that scaffolding through feedback, self-paced learning, and individual learner differences must be considered when designing these systems to enable more effective L2 learning.

Bajorek and Hammond (2019) used a mixed-methods approach in a pre-/post-test study to explore the user experience of American beginner

university students ( $N=134$ ) of French as a second language as they practised for five weeks in ImmerseMe, focusing on pronunciation. The methods used were survey, interview, speech, and backend analytical data. The results revealed that the students increased their speaking practice, feeling equally comfortable (or more so) when speaking to the system as when talking to other human interlocutors in the classroom. The participating students estimated the development of their L2 pronunciation and speaking skills.

Approaching this system as a “serious video game”, Casañ-Pitarch and Gong (2021, p. 219) conducted an experiment with 48 Chinese adult intermediate-level (B1) students living in Spain separated into two groups (one control group) using pre- and post-tests combined with an opinion survey. The results showed progress in four target areas (word-order structure, lexicon grammar, spelling, and content), indicating that ImmerseMe as a complement to other activities could be useful in L2 education.

Soto et al. (2020) have explored the effectiveness of ImmerseMe in improving spoken English among Colombian university students ( $N = 124$ , basic and intermediate levels) during a period of four weeks. Using qualitative data with a quantitative component, focusing on students’ perceptions, produced in a semi-structured survey, interviews, and participant observations during the activity in the system, the study took an exploratory and descriptive approach. The results showed that the students found social interaction within VR and immersive environments to be effective for improving their speaking and listening skills in the classroom. The benefits of VR were that it created meaningful and real-life learning experiences in scenarios that kept the students’ attention while feeling immersed, and interacting with a native speaker.

### 5.3 EXPERIENCES OF L2 SPEAKING WITH CAs

As previously proposed by Reeves and Nass (1996), humans attribute social characteristics to non-human entities based on their behaviour, often referred to as humanisation or anthropomorphism. Previous research on human-machine interactions has studied how humans may interact with ECAs, relating socially, and evolving feelings (e.g., Bickmore & Cas-

sell, 2005; Ebadi & Amini, 2022; Lawson et al., 2021) and establish and maintain a relationship (Bickmore & Rosalind, 2005). However, when the ECAs are too human-like, they might provoke an uncomfortable feeling, known in the literature as the “uncanny valley” sensation, originally found in research with human-like robots (Mori et al., 2012). In their comprehensive multidisciplinary systematic literature review, Van Pinxteren et al. (2020) highlighted that the communicative behaviours of CAs have significantly positive effects on establishing relationships, cooperation, and loyalty among users. According to their typology, the behaviours can be classified into two dimensions: modality (verbal, nonverbal, and appearance) and footing (similarity and responsiveness).

Schroeder et al. (2013) highlighted the importance of considering the design of the voices and gestures of virtual humans for learning. In their design principles for virtual humans in educational technology environments, Craig and Schroeder (2018) added a social component. They suggested that the interface elements, such as human voice, temporal contiguity with speech, gestures, and virtual humans, are social facilitators. In their review, evidence was found for the effectiveness of virtual humans in aiding learning, especially when social elements are added. The importance of social cues, such as facial expression, conversational gestures, speech, and movement, to emulate human-like characteristics and behaviours for creating engaging and credible virtual humans has also been highlighted (Burden & Savin-Baden, 2019) when seeking to create socially immersive experiences and social connections (Divekar et al., 2021). However, investigating the features of virtual humans and their impact on learning, Schroeder et al. (2021) found that students’ perceptions of the type of virtual human voice influenced trust but had little effect on learning. After watching video presentations with two types of voices involved in the virtual humans, the sample of 88 Amazon Mechanical Turk users (female and male adults) in the USA completed a questionnaire based on a revised version of the Agent Persona Instrument. The analysis method was exploratory, using a combination of unsupervised machine learning (k-means clustering) and non-parametric statistical analyses.

Interaction between humans and CAs for L2 learning by means of conversational AI is not a new research interest. However, until recently, it seems to have focused mostly on the effectiveness of CAs, and not on their



design. In 2002, using “agents and bots for CALL” was mentioned as a promising and maturing function (Schwienhorst, 2002, p. 202). Recently, a research team presented preliminary evidence that a CA’s (or system’s) nonverbal emotional feedback through facial expression increased students’ willingness to communicate in L2 (Ayedoun & Tokumaru, 2022). The study was grounded in social agency theory (Atkinson et al., 2005), using verbal and visual cues as humans do to encourage social interaction and an experienced partnership in human–computer interaction. Ayedoun and Tokumaru (2022) enabled highly realistic animated virtual humans with believable nonverbal listening behaviours (facial expressions, nodding, and gaze) and let 12 Japanese undergraduate university students evaluate the meaningfulness of a prototype system through a Wizard of Oz (Woz) experimental style. Data was generated in a preference questionnaire, interaction time logs, and transcripts of the five-minute speaking session for a combined descriptive analysis. Ayedoun and Tokumaru (2022) welcome critical reflection on how to design CAs with emotional expression and empathic behaviours to enable efficient L2 learning for students in terms of building human-to-CA (technology) relationships through social interaction as a complement to human-to-human relationships in interaction. This aligns with the work of Walker and Ogan (2016), who early on envisioned the importance of carefully designing AIED systems, e.g., SDSs and ECAs, to embed social relationships in interaction for an attentive and personalised beneficial learning experience.

Timpe-Laughlin et al. (2022) have suggested that future L2 research with SDSs should focus on the impact of the ECA’s facial expressions and body language because participating students ( $N = 51$ ) stated that they missed reactions, emotions, and facial expressions in the ECAs compared to human-to-human interaction. However, in their meta-analysis of SDSs, Bibauw et al. (2022) found no significant differences in L2 development in relation to the type of embodied CA (virtual human or robot).

Studies exploring L2 practice with CAs in SDSs have found them helpful in providing feedback. However, students have also reported feeling frustrated when the CA does not understand them properly due to communication breakdown, interactional design, constraints, or technical issues (e.g., ASR; Fryer et al., 2019; Johnson, 2019a; Gallacher et al., 2018; Wang et al., 2022). Alm and Nkomo (2020) used thematic analy-

sis and sentiment analysis to conduct a qualitative analysis of the student (user) experience data generated from various online social media network platforms, where the purposive sample discussed their experiences of using chatbots for L2 learning. The results indicated that interaction with chatbots was an enjoyable way of practising L2, despite their limitations. According to Bibauw et al. (2019):

To design a dialogue-based CALL application is to find an adequate balance between constraints, which guide and focus the user production, to reduce its unpredictability and allow its automated processing, and freedom left to the learner to express their own meanings interactively. (p. 37)

## 5.4 RELEVANCE TO THIS THESIS

Overall, as previous research has shown, teachers are important for the use of digital tools in general in language education, and for the two SDSs selected for this thesis in particular, in terms of aspects such as ensuring meaningful, contextualised L2 practice of speaking and listening. Teachers have generally expressed positive views and experiences of implementing digital tools, although they acknowledge some barriers to employing them, such as being time-consuming and teachers' own lack of adequate digital competence and training in using them. Previous studies have not explored how digital tools are experienced in the Swedish L2 education in relation to teaching and learning L2 speaking skills.

As shown in the related research, prior studies have presented a scattered and dispersed field, with a lack of independent studies based on students' self-reported experiences in a school context with students younger than university level, or in Europe. There have been a number of separate small exploratory studies employing a mixed-methods approach over a short period of time, mostly with English as the target language. Logbooks for producing student data over time were not a frequently used method. Most previous studies have focused on measuring effects and have analysed quantitative data, but often without achieving statistical significance due to small sample sizes and short trials. The results have highlighted a

motivational boost in students' confidence in speaking and reduced anxiety, which has a positive effect on their willingness to communicate in L2.

Based on the established key SLA principles, it is central for this thesis that students conceptualise the L2 by engaging in contextualised social interactions with instant feedback in a safe environment. Such experiences are recognised as being facilitated in SDSs with ECAs that enable dialogue to various extents, depending on the constraints of the system. There is established positive evidence for the effectiveness of conversational AI in L2 education, with a stronger impact on low- or intermediate-proficiency students (beginners) than on high-proficiency students. In studies employing the same SDSs as in this thesis, Enskill studies reported on analyses of performance in a good, fun, and easy practice generating increased confidence in speaking. ImmerseMe studies focused on improved pronunciation and increased time spent speaking the target language.

Prior studies of experiences of speaking with CAs and ECAs from related fields, such as virtual human research, have shown that their capabilities demonstrate benefits for the social aspect of generating social agency and enabling social interaction, which in turn have shown relevant for learning. In DB-CALL, this seems to be an understudied area, and this thesis will extend previous results and fill the gap in relation to L2 development and learning by engaging in speaking practice in SDSs. The next chapter presents the theoretical underpinnings for analysing the data produced in the studies that make up this thesis.



## CHAPTER 6

# THEORETICAL UNDERPINNINGS

This chapter introduces the central concepts employed in this thesis, which are grounded in communication and the sociocultural ideas of L2 learning and development through social interaction (e.g., Vygotsky, 1978) and an emphasis on active learning (Dewey, 1959). These concepts are recognised within both SLA and CALL. The educational activity is studied in line with Long's Interaction Hypothesis (1996), stressing the crucial role of interaction in L2 learning and development. Finally, this chapter addresses the central concept of experience in relation to the framework Relationship of Inquiry (RoI; Cleveland-Innes et al., 2014; Stenbom, 2015) and its operationalisation in the iteratively developed framework FoSCAI to refine the conceptualisation of experience as a methodological and analytical lens. FoSCAI shares the same theoretical underpinnings from Vygotsky (1978) and Dewey (1959) as its framework of inspiration (RoI), which also investigates student-activity-based online learning through interaction, but in human-to-human dialogues.

### 6.1 COMMUNICATION AND SOCIAL INTERACTION

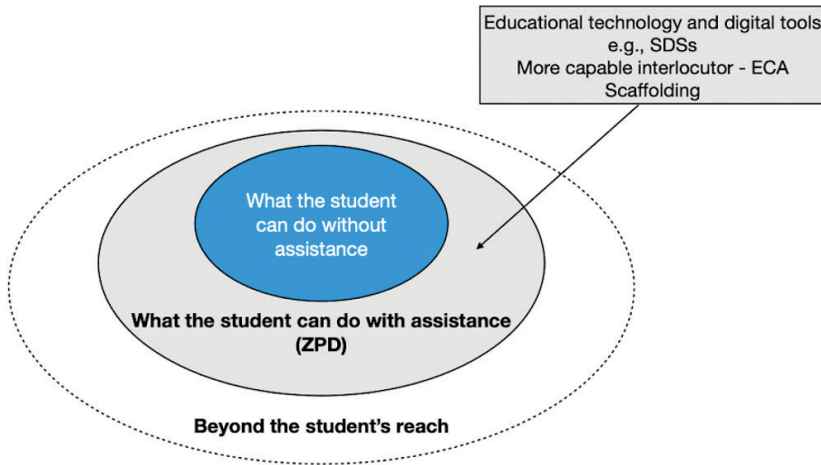
Communication is an invaluable part of our lives. We use it to create bonds with other people, and being able to speak with others is often the reason why students learn a new language (Walker & White, 2017). Communication encompasses verbal modes in aspects such as speech and syntax, as well as nonverbal modes, including gestures, body language, and social relations. Vygotsky (1978) conceptualised communication as a fundamental sociocultural process that facilitates learning and cognitive devel-

opment. Hence, communication plays a vital role in mediating knowledge and meaning within social interactions, supporting language learning, and contributing to the formation of higher mental functions. Growing up in an environment where a particular language is used actively helps an individual to naturally learn and develop that language, including the cultural aspects associated with it.

Vygotsky's ideas hold particular significance in the context of L2 learning (Lantolf et al., 2014; Mahn & Fazlehaq, 2020). Language learning is viewed as a social and cultural process, with communication acting as a bridge enabling students to engage with the target language and culture. Through social interaction and communication with a more knowledgeable speaker, students can access the Zone of Proximal Development (ZPD; Vygotsky, 1978), to advance their potential level of competence; for instance, in speaking and cultural understanding (Lantolf et al., 2014). Ohta (1995) expressed ZPD in an L2 context as “the difference between the L2 learner's developmental level as determined by independent language use and the higher level of potential development as determined by how language is used in collaboration with a more capable interlocutor” (p. 96). This emphasises the importance of social interaction in language learning within the ZPD. It is recognised that students can achieve more when engaging in optimally challenging activities and within reach of assistance (Gibbons, 2015). This temporary assistance is referred to as scaffolding, and was introduced by Wood et al. (1976). In an educational context, it can refer to the support provided by a more knowledgeable teacher or peer or by digital tools (e.g., SDSs with ECAs), to help students achieve their goals. Gibbons (2015) described it as: “a special kind of help that assists learners in moving toward new skills, concepts, or levels of understanding” (p. 16). As illustrated in Figure 6.1, digital tools as mediational tools are artefacts that can enhance cognitive abilities, provide additional practice, and offer support to students (Säljö, 2022). The role of the more capable speaker is to provide adequate support to help the student move towards more advanced skills and achieve their desired outcomes.

**Figure 6.1**

*Zone of proximal development and scaffolding*



Scaffolding can also include breaking down complex tasks into smaller steps, providing support, feedback, and various resources, for then gradually reducing the support as the student becomes more able to speak independently. Students must be provided with the “necessary scaffolding, input, and guidance” (Goh & Burns, 2012, p. 6) in relation to their speaking activities and not just left alone to struggle with the development of their speaking skills. However, there is also a risk of the student compensating for his or her weaknesses and delaying development and learning due to the overuse of scaffolding (VanPatten & Benati, 2015, p. 180). As students engage in social interactions and receive scaffolded support, they may gradually internalise language knowledge and communicative strategies, incorporating them into their internal individual mental processes. This will lead to more independent and flexible L2 use (Lantof et al., 2014; Vygotsky, 1978).

In this thesis, these theoretical underpinnings imply that L2 learning and development are enabled when students engage in social interaction through spoken dialogues in sociocultural environments. However, in the studied systems, the other interlocutor is represented by an ECA as

a native speaker of the target language (the more capable partner), enabling face-to-face interaction with instant conversational feedback. Hence, interactional activities are central to L2 learning and development when it comes to developing speaking skills (Gass & Mackey, 2015; Lantolf et al., 2014; Loewen & Sato, 2018; Long, 1996).

### 6.1.1 INTERACTION HYPOTHESIS

According to Long's Interaction Hypothesis (1996), comprehensible input and output, in combination with appropriate feedback, facilitate language development. Gass and Mackey (2015) have also highlighted the importance of input, which for L2 spoken development refers to utterances in the target language that the student has heard and understood (Flege, 2009). When comparing spoken input with written input, the aspects highlighted include turn-taking and recasting of the spoken utterances, in contrast to written sentences without recasting (Goh & Burns, 2012; Nilsson, 2018). Output is the utterance produced by the student. When the meaning is not communicated or understood as expected, the student negotiates by using other resources to establish a mutual understanding in the dialogue around the task that needs to be solved (Ellis et al., 2020). The student may have to reformulate the utterance with some output more like the target language in terms of accuracy (Gass & Mackey, 2015).

When learning L2, feedback is crucial. It can be either explicit, through correction of the spoken language, or implicit through, for instance, negotiation strategies, confirmation, requests for clarification, or recasts when rephrasing utterances during conversations (Gass & Mackey, 2015; Van-Patten & Benati, 2015). In a classroom context, "input-feedback-modified output in oral interaction with competent speakers" (Goh & Burns, 2012, p.19) is often limited due to practical and physical constraints, which also make it difficult to provide feedback to each student after speaking activities. In these cases, educational technology might become a possible facilitator (Li, 2017).

In this thesis, in student-to-CA interactions, both speakers must act as if they understand each other in the target language and work to negotiate meaning, which is enabled to a varied extent following the system's interactional design. In line with the Interaction Hypothesis (Long, 1996),



implicit feedback occurs naturally in the interaction through, for instance, confirmation checks and recasts if these are needed to proceed, but without disturbing the speech production in the interaction. Additionally, in an SDS, there is the possibility of receiving summarised feedback on the performed utterances after the simulated conversation and, to some extent, corrective feedback in terms of simultaneously produced written transcripts of the interaction.

## 6.2 LEARNING BY SPEAKING THE TARGET LANGUAGE

The socio-cultural theoretical underpinnings of this thesis are combined with ideas about active learning and problem-solving. In line with Dewey's (1959) emphasis on the importance of engaging in hands-on activities in an enjoyable and meaningful way close to the actual skill to be learnt and developed, L2 speaking skills are practised in interaction in simulated conversational scenarios. The aim of the speaking activities is for students to be able to interact independently with other humans in everyday-life situations. Dewey argued that learning is most effective when it is connected to the student's experiences and interests, and involves engagement with the subject of the interaction. These ideas are recognised in task-based learning, which highlights the use of tasks that seem authentic and relevant to the students in terms of their everyday lives and interests (Ellis, et al., 2020). Although both Vygotsky's (1978) sociocultural and Dewey's (1959) pragmatic ideas emphasised learning and development as the construction of knowledge, they also highlighted the role of the teacher and the importance of tools in supporting the process. According to Collins and Halverson (2010):

Information technologies have pushed us to a radical, learner-oriented understanding of development and learning. Information technologies foster a more hands-on, activity-based education. Computers are highly interactive and provide a variety of systems to accomplish meaningful tasks. Hence, they are more aligned with the "learning by doing" view of education than with the "absorption of cultural knowledge" view of education that permeates schooling. (p. 20)

As explored in this thesis, older ways of teaching and learning L2 speaking skills are now being challenged by new technologies such as conversational AI, where students can act and regulate their own practice of L2 speaking with ECAs and experience another learning situation.

### 6.3 EXPERIENCE

The concept of *experience* is central in this thesis. Both Dewey and Vygotsky emphasised the role of experience in relation to learning and development. Vygotsky (1978) highlighted the importance of social and cultural experiences in shaping cognitive processes, while Dewey (1959) emphasised the significance of experiential learning, in which students actively engage with their environment and reflect upon their experiences in order to construct knowledge and meaningful learning. In their view, the complex concept of experience is an indivisible unit, which includes the people, environment, mutual effects on each other in social relations, and affect across time and space (Roth & Jornet, 2014). As explained by Vygotsky (1978) when talking about experiences, the Russian word *perezhivanie* integrates not only outer social environmental aspects but also inner intellectual and affective aspects. In more recent years, these have also been found to have a bearing on L2 learning (Lantolf & Swain, 2020). A more contemporary interpretation of experiences per se could be seen as differentiating humans [we] from machines:

We do have experiences, feel and have emotions. And we do learn from and through our feelings. At the heart of this process, therefore, is the learner as a whole person and, therefore, the learner's self. But self is formed through existing and interacting with people. We actually learn to become a person, and this occurs within our life-world; it is a social process. (Jarvis, 2006, p. 6)

In this thesis, the student-to-CA interaction forms a new kind of social interaction, occurring in or through an SDS, which challenges traditional ways of studying social interaction in relation to learning. Experience is a broad and complex concept, which includes various interrelated aspects; therefore there was a growing need during the studies conducted for this

thesis to further operationalise experience by dividing it into various dimensions for the organisation and analysis of data. This is further elaborated upon in the next section.

## 6.4 FRAMEWORK OF STUDENT–CA INTERACTION

The FoSCAI framework was designed to refine the exploration of students' experiences within an SDS (Ericsson & Jansson, 2021). It was used as a methodological and analytical lens, and then applied in Study 4. FoSCAI would not have been developed in the same way without the empirical and theoretical insights gained from Studies 2 and 3. Theoretical and methodological similarities were found in two already-established frameworks used for research into online and blended learning, the Community of Inquiry (CoI; Andersson et al., 2001) and the Relationship of Inquiry (RoI), being an adaptation of CoI for one-to-one online coaching (Cleveland-Innes et al., 2014; Stenbom, 2015). The theoretical underpinnings of CoI, RoI, and FoSCAI were contained in Vygotsky's (1978) sociocultural ideas, which emphasise the role of social interaction with a more knowledgeable character in the ZPD, combined with Dewey's (1959) ideas of learning by active problem-solving in dialogue, through social interaction shaping meaningful learning experiences. Next, follows a brief presentation of *presence*, *experience*, and *interaction* as used in CoI and RoI, and how they have informed and inspired the development of FoSCAI, with its four dimensions of experience (cognitive, emotional, social, and teaching) of students' interactions with ECAs.

### 6.4.1 PRESENCE, EXPERIENCE, AND INTERACTION

In previous studies CoI has been employed to study presence, referring to the perceived connectedness and engagement of teachers, students, and their learning environment. It encompasses the interrelated elements of social presence, cognitive presence, and teaching presence, which collectively contribute to the educational experience and the facilitation of meaningful online or blended learning interactions (Garrison et al., 2000). The CoI framework has primarily been used to study and design online learning in higher education by, for example, analysing written human-

to-human, computer-mediated communication for collaborative learning and coaching (Andersson et al., 2001). The use of CoI has just recently been explored outside human-to-human interactions in Wang et al. (2022) investigating learners' perceived presence in AI-supported language learning via the students' reported experiences.

As a further adaption of CoI, RoI is used in one-to-one coaching online in subjects such as mathematics (Cleveland-Innes et al., 2014; Stenbom, 2015). RoI includes a fourth element, emotional presence, which refers to expressed emotions that extend beyond the social presence, such as activity emotion, outcome emotion, and feelings in relation to the technology, content, student, and teacher. A coach or tutor can be a teacher, and so can an older student who knows more and supports the inquiry activity to facilitate the learning process of the tutee, the student. Stenbom et al. (2016) defines an educational relationship of inquiry as "a conceptual connection that is built between two persons, here a coachee and a coach that engage in critical discourse" (p. 38) to solve an educational task in a one-to-one online setting.

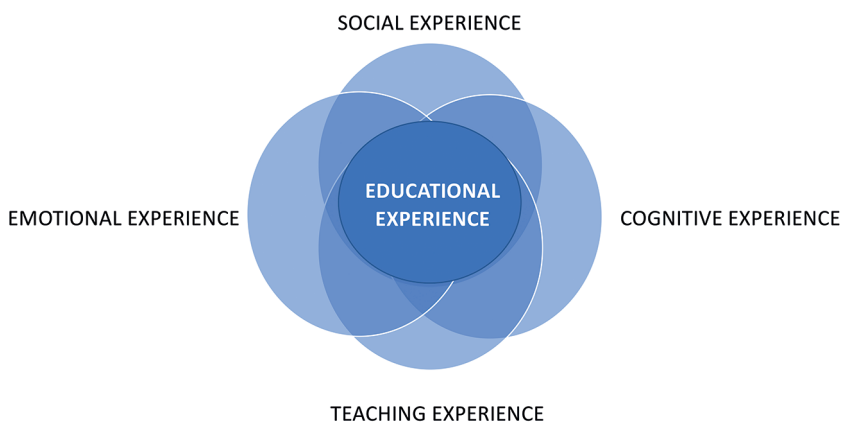
However, in this thesis, the one-to-one conditions in the interaction were different, and the focus was not on two human participants but on one human and one ECA. The student interacted orally with an ECA in order to solve a task in the target language, rather than interacting about learning mathematics with a human being, as in RoI. Hence, there were no mutual experiences such as those that occur in human-to-human interactions. The aspects of experience, such as behaviour, thoughts, feelings, and views were not mutually transforming (Dewey & Bentley, 1999), because the ECAs within the SDS are not human. The focus in Study 4 was, therefore, on the *students' experiences*. There was no human being in the coach's role in this learning situation, as in the original idea of RoI, which focuses on presence.

The reason for building on RoI, with its four elements for operationalising, in this thesis was that emotions are considered to relate to cognition in L2 learning and development (Lantolf & Swain, 2020; Swain, 2013). This motivated an exploration of the students' experienced emotions, both during and after practising their L2 speaking skills. Consequently, the RoI framework's four elements were used as inspiration for the designed framework, FoSCAI.

The operationalisation concerned students' experiences of practising L2 speaking in the SDS, which were dimensionalised into cognitive, emotional, social, and teaching experiences (see Figure 6.2) as a novel analytical construction, due to the complexity of experience (Roth & Jornet, 2014). The figure below (Fig. 6.2) is a visualisation of this analytical construction. All four dimensions overlap in the middle, thus together forming the educational experience, without needing to know the precise size of each dimension. The interrelated dimensions are studied both separately and collapsed together. However, the overlapping regions demonstrating relationships between the dimensions are not the focus of this thesis.

### Figure 6.2

*Visualisation of the four dimensions of experience, all interrelated in the student's educational experience of practising L2 speaking in an SDS*



*Note.* The figure does not claim to be proportionally correct.

In terms of what RoI and FoSCAI mean for how to organise teaching, there are similarities between the frameworks. FoSCAI also takes a student-oriented approach, with some guidance and scaffolding from the whole SDS (coach), including the ECA, to facilitate learning, which here refers to the development of L2 speaking skills (e.g., interaction). The whole SDS is understood as being similar to a coach/tutor in that it provides

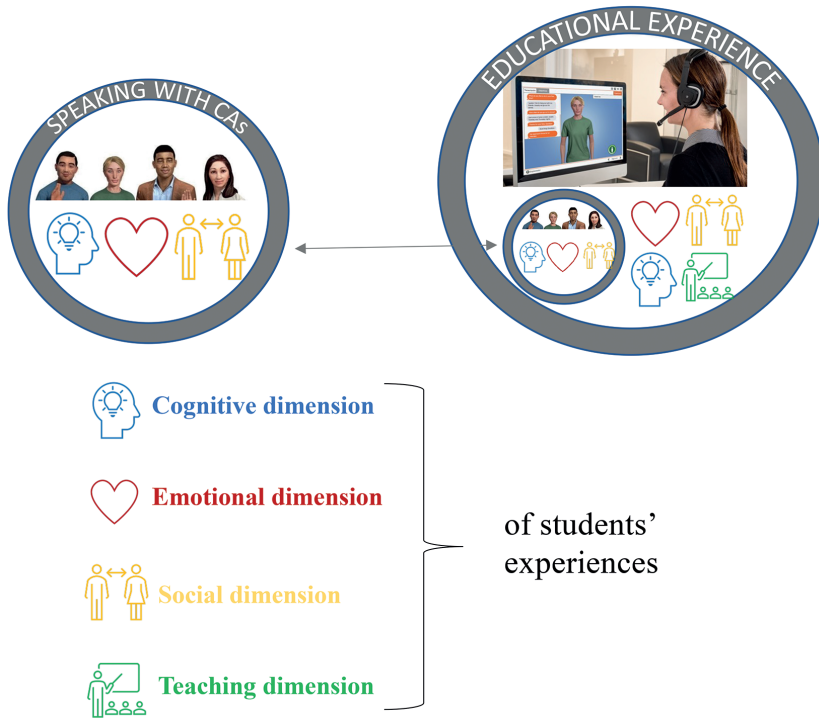
scaffolding possibilities. It does so through supportive features and formative individualised assessment (Ericsson & Jansson, 2021) to effectively support the student within his or her ZPD to nurture the development of speaking skills. FoSCAI suggests that the student in an SDS is “coached” in the dialogue through the ECA’s instant feedback during the interaction in the selected SDS, and in the turn-taking between the interlocutors. As the name suggests, FoSCAI is applicable to any type of CA, not only the embodied CAs employed in this thesis. Hence, the term CA is used in relation to descriptions of FoSCAI.

#### 6.4.2 DIMENSIONALISING OF EXPERIENCE

The educational experience of the students concerns what it is like to *practise L2 speaking in an SDS* (see Figure 6.3) in terms of the development of L2 speaking skills. These experiences are dimensionalised into four inter-related dimensions: cognitive, emotional, social, and teaching. The *L2 speaking experience with a CA* is separated for analytical purposes and only dimensionalised into cognitive, emotional, and social dimensions, before then also being collapsed together and included in the educational experience. A coloured symbol illustrates each dimension of experience in line with FoSCAI when visualising the framework and the outline of the four dimensions.

**Figure 6.3**

*Visualisation of the Framework of Student-CA Interaction (FoSCAI)*



*Note.* To study the learning activity of practising L2 speaking skills in an SDS, FoSCAI allows for both individual analyses of each dimension and an analysis of the learning activity as a whole including all four dimensions.

Next, follows a description of how the dimensions of FoSCAI were defined and adapted:



The **cognitive dimension** concerns the mental ability to maintain a sustained interaction with the CA, which is initially triggered by being given a task to solve at the beginning of the session. In the educational context explored in this thesis, the students are supposed to be able to interact in L2 with fluency, make themselves understood, and understand the CA in the simulated everyday-life situation, while being challenged and supported, and still maintaining the dialogue. Information is exchanged between the CA (SDS) and the student. The result depends upon the student's interpretation of the various pieces of information and utterances of the CA, used together with earlier knowledge gained during the dialogue, as well as the CA's active participation in the dialogue and interpretation of the student's utterances. This dimension also concerns what the students estimated having practised during the learning activity (e.g., speaking, listening, reading, and writing), which speaking skills (e.g., pronunciation or interaction), how they solved any emerging problems in the communication, and the extent to which the tasks (learning objectives) and dialogues were solved.



The **emotional dimension** is closely related to the social dimension. This dimension concerns emotions in relation to the interaction and judgements about the L2 speaking activity, the outcome emotion related to the results of the activity, and the emotions directed towards the CAs. The indicators analysed concern the students' feelings during the speaking practice (e.g., frustrated), within the environment (e.g., safe), about the learning outcome, and hence the achieved learning objectives (e.g., satisfied), how well prepared the students felt when using L2 in everyday-life situations, in their real-life interactions outside the institutional, educational context. Furthermore, this dimension includes how it felt to speak with a CA (e.g., fun, safe, anxious, exciting, meaningful, or stressful), the outcome emotion (e.g., feeling satisfied after the conversation), and feelings during the interaction with the CA (e.g., happy, sad, engaged, disengaged, frustrated, etc.).





The **social dimension** concerns relationship cohesion, the extent to which the student humanises the CAs, relates to them socially, and/or experiences collaboration during the interaction. This dimension also involves the extent to which the students feel immersed in the conversational scenarios and act as if they were physically present with the interlocutor, and as if the conversation was happening in real life. Furthermore, the social dimension focuses on experiences of feeling comfortable in the interaction with the CAs, and speaking L2 in this way compared to speaking it with someone in the classroom. Also, this dimension includes how the student experienced the signals displayed by the CA during the conversation (e.g., engaged, disengaged, happy, sad, alive, deadpan). This dimension is adapted to this particular situation, with the interaction between a student and a CA online without a real-life person as a coach, as is the case in RoI. There are constraints on the student-CA interaction, and therefore RoI's category about open communication (e.g., Stenbom, 2015, p. 12) was not fully applicable, with no opportunity for an open human-to-human conversation online.



The **teaching dimension** concerns the pedagogical design and organisation of the SDS, including such aspects as learning goals, learning methods, facilitating L2 practice (supporting and monitoring the process), and direct instruction to stimulate the cognitive and social process for realising meaningful L2 practice to stimulate learning and development. In FoSCAI, the teaching dimension consists of three parts: (i) the CA as interlocutor (giving instant feedback), (ii) the whole SDS, including its supportive features and an eventual dashboard monitoring information, and (iii) the student regulating his or her practice in the system through making choices and using the supportive features (i.e., transcripts, objectives, and direction and monitoring information).

In this thesis, in line with FoSCAI, the teaching dimension was only applicable when exploring the experiences of *practising* L2 speaking in an SDS (RQ2), and not when specifically exploring the experiences of L2 speaking with ECAs (RQ3). However, speaking with an ECA is the central part of the learning activity. The indicators analysed included, for instance, the relevance of the content of the scenarios and the usefulness of the practice as a learning activity for developing L2 speaking skills in terms of communicating with an ECA and receiving instant feedback during the interaction.

### 6.4.3 EDUCATIONAL EXPERIENCE

In this thesis, the conceptualisation of students' *educational experience* was refined iteratively during Studies 2–4 based on an empirical and theoretical understanding applied via FoSCAI (see Figure 6.3). In Studies 2 and 3 (before the development of FoSCAI), there was a more general overall picture of the experience of practising L2 speaking skills in the SDSs. In Study 4, *educational experience* was conceptualised in line with the operationalisation into the FoSCAI framework, by applying the four dimensions including the central aspect of L2 speaking with CAs (dimensionalised into cognitive, emotional, and social dimensions). The definition of educational experience is hence the comprehensive experience of all four dimensions together of practising L2 speaking skills in an SDS, with the aim of being beneficial for L2 learning and development in terms of speaking skills. When students themselves rated their own overall educational experience, this was referred to as “overall experience” of practising in an SDS in the analysed indicators, as further explained in Chapter 7, which describes the methods applied.

## CHAPTER 7

# RESEARCH DESIGN AND METHODS

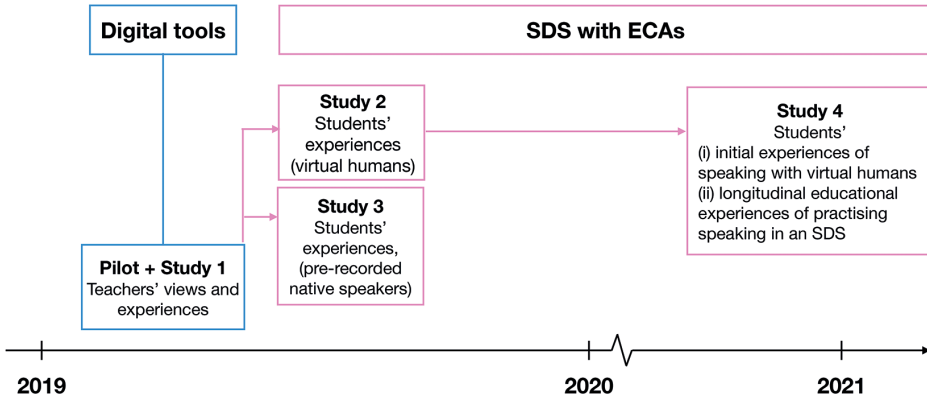
In the first section of this chapter, the research design of this thesis is addressed. The first section includes the motivation for using experience as an analytical unit, a presentation of the participants, the set-up of the studies, the selection of SDSs, and an overview of the chosen systems (Enskill in Studies 2 and 4, and ImmerseMe in Study 3). In the second section, there is a presentation of the data produced, the methods used for gathering this data, and a description of each study, including the methods used, before concluding with a discussion of the methods of data analysis and ethical concerns.

### 7.1 RESEARCH DESIGN

An overview of the research design and timeline of the thesis' four studies are presented in Figure 7.1. The whole thesis combines qualitative and quantitative analyses. Primarily qualitative data was used in Studies 1–3, and quantitative in Study 4.

**Figure 7.1**

*The research design and the timeline of the four studies*



To address RQ1, about language teachers' views and experiences of using educational digital tools in their teaching, with a particular focus on speaking skills, first a pilot study was conducted, and then Study 1. As a springboard for establishing the further direction of this thesis, before conducting studies with students and introducing SDS, the Pilot and Study 1 were completed to enable an understanding of teachers' perspectives and the current Swedish L2 educational situation in relation to digitalisation and speaking skills. Since teachers are responsible for designing lessons to aid in students' learning and development of speaking skills, including the choice and implementation of educational digital tools, it was regarded as a useful first step to determine whether there was any need for and/or interest in this kind of conversational AI as a complement to today's teaching methods. Moreover, the studies with students depended on the willingness of the teacher to actively participate with his or her students by using selected SDSs.

In this thesis, the class teacher is not actively involved in the learning situation in the SDS. Instead, the teaching is divided between the instant feedback provided by the ECA, the whole SDS scaffolding with its supportive features, sometimes giving corrective feedback and assessment, and

the student herself or himself, by self-regulating their learning in the system. To address RQ2 and RQ3, three studies with students (Studies 2–4) were conducted to explore how lower-secondary students experienced practising L2 speaking skills in an SDS and speaking with ECAs.

Studies 2 and 3 were designed in similar ways, using two different SDSs to provide L2 speaking practice so as not to depend on only one system. The differences in the systems' design provided broader insights into students' experiences of speaking with two types of ECAs – virtual humans and pre-filmed native speakers – and using English as L2 and German as L3 for the target languages. Study 2 was labelled *iSpeak\_Ge*, and Study 3 was labelled *iSpeak\_Eng* to distinguish them from each other based on the target language, and signalling that it is about speaking. These labels will henceforth be used in the tables.

Based on Studies 2 and 3, there was an identified need for a more systematic study of the students' experiences, including over time. Hence, the framework FoSCAI was developed (see Section 6.4.) for organising and analysing the data produced in Study 4. This study had a longitudinal design, involving repeated speaking sessions in the SDS over time, and was labelled *iSpeakL\_Eng*. This label will be used in tables to facilitate distinguishing the studies for the reader.

In addition, subgroups were introduced into the analyses due to important individual differences in the student experiences in Studies 2 and 4. To identify patterns, two subgroups were introduced, based on (i) gender (female and male) in line with the guidelines suggested by Heidari et al. (2016) and (ii) proficiency level (high or low, based on the teacher's assessment of speaking), as background variables as used in DB-CALL and SLA research (Bibauw et al., 2019). However, since neither gender nor proficiency-related questions were explicitly the main aim of this thesis, these results could be seen as foundations for further work and are not further investigated here.

### 7.1.1 EXPERIENCE AS AN ANALYTICAL UNIT

When introducing CALL to students, as suggested by Levy (2015), it is critical to understand their experiences and “how the participants experience and interact with a phenomenon at a given point in time and in a

particular context...” (Heigham & Croker, 2009, p.7). In this thesis, the focus is on understanding the students’ subjective comprehensive experiences of practising L2 speaking in an SDS. The students self-reported what they experienced, referring to many aspects of those experiences, such as the speaking activity with ECAs, reflections, thoughts, feelings, and more detailed views and evaluations of what they had actually practised in the SDS in relation to L2 learning and development. The students estimated what emotional experiences had been generated for them during the speaking practice in the SDS, how they felt about the ECAs, and the apparent emotions displayed by the ECAs in the interaction (labelled signals to avoid giving them human associations). Underlying emotional processes – such as enjoyment, excitement, or boredom, to mention just a few – may consequently play an important role when practising speaking in an SDS, as well as the feelings related to the ECA, the ability to socialise while practising, and various other aspects of practising speaking in an SDS. In sum, the empirical material was built on self-reports, with the students reporting what they did, thought, felt, learnt, and developed.

Experience covers similar aspects to the educational experience that was iteratively conceptualised in FoSCAI, which enabled the delineation of experience into four dimensions. This enabled deeper insights into the underlying aspects of the reported experiences, while also generating an aggregated, comprehensive understanding of the students’ educational experience of practising L2 speaking. When the students themselves rated their overall educational experience of practising L2 in the system, this was referred to as the overall experience. For example, there was one post-trial questionnaire item asking about their overall experience in the SDS, and in Study 4, post-trial emoji rating in the digital logbook LoopMe (2023, Lackéus, 2020)

### 7.1.2 PARTICIPANTS

In Study 1, the participating teachers were a convenience sample of Spanish teachers ( $N = 9$ ) in the pilot and L2 teachers ( $N = 36$ ). In total, there were 45 teachers, four male and 41 female. They taught languages professionally at various levels, with a clear majority teaching at lower-secondary level and in English. This research’s participating teachers represent the

most common non-probability sampling used in L2 research, convenience sampling (Dörnyei & Taguchi, 2010). The teachers participating in this thesis were selected due to being suitable for the purpose of the study, geographical proximity, and ease of accessibility.

The participating students ( $N = 86$ ) were from the same school in the southwest of Sweden, selected through one teacher who volunteered from Study 1 to continue participating in Studies 2–4 with three groups of students. Hence, the recruited students were a convenience sample consisting of 36 female and 50 male students in 7th or 8th grade, studying English (L2) and/or German (L3). Two male students participated in both Studies 2 and 3. Three students chose not to participate in Study 2, two failed to complete the data on all occasions, and one student changed schools during the longitudinal study.

### 7.1.3 SETTING AND REALISATION OF STUDIES 2–4

The empirical setting consisted of Swedish lower-secondary school classrooms, with groups of students simultaneously engaging in individual practice. The learning activity was practising L2 speaking skills in an SDS. As illustrated in Figure 7.2, the learning situation included the student and the virtual environment in the SDS. Although everything took place in an educational context, primarily set in the classroom, the surrounding setting within the room was not included in the analysis. In Study 4, the students conducted a few speaking sessions at home due to the COVID-19 pandemic restrictions.

## Figure 7.2

*Example of the learning situation: the student and the SDS*



*Note.* ©Alelo Inc. Used with permission. Private photo.

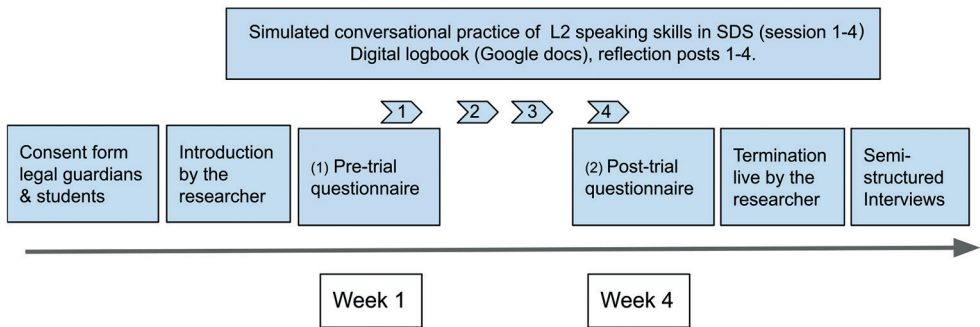
In Studies 2–4, each student was already equipped with a Chromebook from school and was offered the loan of headphones with a microphone. Each student received an account with an anonymised log-in to the SDS (Enskill or ImmerseMe) and logged in via their individual Google accounts to a digital logbook. In each speaking session, the students practised speaking German or English in everyday-life situations simulated in the SDS for approximately 15 minutes, followed by writing reflections in their personal digital logbook. In Studies 2 and 3, they reflected in a shared document, and in Study 4, in LoopMe. The students answered questionnaires individually in Google Forms on their Chromebooks. Semi-structured interviews were conducted with some of the students post-trial and with the teacher and a trainee teacher both pre- and post-trial. The teacher had no access to the students' logbooks or questionnaires in order not to intrude upon the students' privacy when answering the study, and to allay any fears of it affecting future L2 assessments.



The teacher, licenced in English and German, expressed interest in participating in the three studies, and their school's principal consented, as did the students and their legal guardians. The teacher and the researcher (myself) decided on the time interval and length of studies in order to fit with other planned educational activities combined with the research aims. The overall realisation processes of Studies 2–4 are displayed in Figures 7.3 and 7.4 and more detail is provided in section 7.3, including Table 7.2.

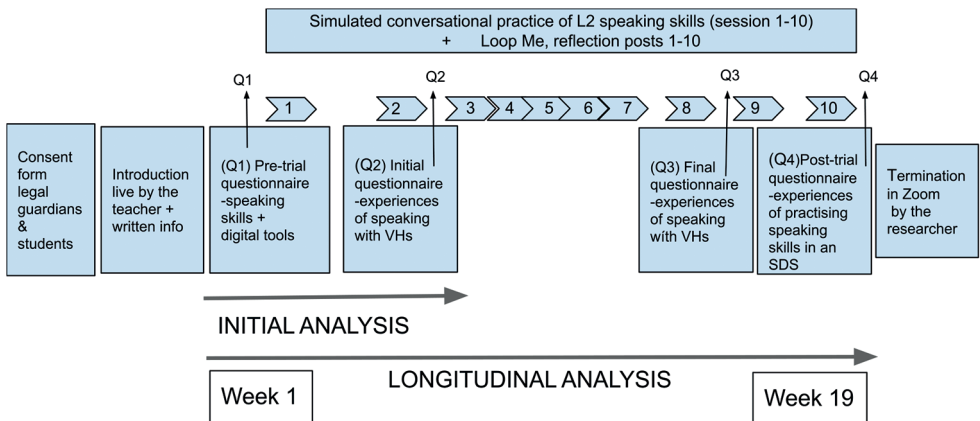
**Figure 7.3**

*The realisation of Studies 2 and 3*



**Figure 7.4**

*The realisation of Study 4 over time*



*Note.* Q = Questionnaire (1–4), pre-and post-trial questionnaires

#### 7.1.4 SELECTION OF SPOKEN DIALOGUE SYSTEMS

The students practiced speaking skills in two different SDSs, Enskill (Alelo, 2022; Studies 2 and 4) and ImmerseMe (n.d.; Study 3). I decided on the selection criteria, which concerned comparable aspects; both were low-immersive (Dhimolea et al., 2022), hence desktop versions of the SDS. The students had to be able to practise L2 speaking skills in conversational simulations, face-to-face, spontaneously and with natural speech. The conversations had to occur in everyday-life scenarios with a human-like interlocutor (ECA) within a system that registers, gives instant feedback, and, if possible, also analyses the conversation to give the student individual assessment. The system may measure and collect data produced based on the student's performance in the conversational scenarios for purposes of understanding and optimising learning and the surrounding environment. Additionally, the systems had to be publicly available. It was hard to find publicly available SDSs fulfilling the set criteria when searching and testing available digital tools for L2 learning and development.

Among the 13 pedagogical resources and digital tools tested<sup>1</sup> in autumn 2018, the selected systems Enskill and ImmerseMe were, at that point, the only two available that fulfilled the set criteria. These systems were identified as goal-oriented contextualised dialogue systems, according to Bibauw et al.'s framework (2019). In Enskill, the target language was American English, with few constraints in the conversation, enabling spontaneous speech. ImmerseMe included nine languages at that time, but with more constrained practice, which limited the student's speech.

The two selected systems were first tested on teachers in the pilot study. Agreements were reached with both companies behind the selected SDSs making it possible to use the systems in the studies. Next there follows a description of the two systems in terms of their interrelated technological, interactional (system and dialogue type including constraints), and instructional design characteristics, along with the implications of these characteristics for pedagogical aspects of the practice using them.

---

<sup>1</sup>Tested pedagogical resources: Digilär and NE. Tested educational digital tools: Babbel, Busuu, Duolingo, Enskill, ImmerseMe, Indian gogo, LingQ, Memrise, Mondly + Mondly VR, and WordDive.

### 7.1.4.1 ENSKILL

Enskill was used in Studies 2 and 4. The system is designed to enhance speaking skills, primarily spoken interaction, through asynchronous simulated everyday-life conversational interactions with ECAs. The target language is American English on A1 and A2 proficiency levels with their descriptors for online interaction, according to CEFR (Council of Europe, 2020). Enskill was initially designed for higher education by an American company on the same platform as its earlier system which was used by the military for training in interactional and cultural skills (Johnson & Valente, 2009). The dialogues take place in scenarios within an animated VR-based environment, with can-do definitions and an action-oriented approach. Every simulation has several learning objectives for the students to complete in order to solve the task; for instance, ordering at a restaurant giving directions or buying a ticket (see Figure 7.5).

**Figure 7.5**

*Simulation of buying a train ticket to New York in Enskill*

The screenshot displays the Enskill English CEFR A2 Demo interface. At the top, the logo "alelo" is on the left, and "Enskill English CEFR A2 Demo" is in the center. On the right, there are icons for a menu, help, volume, and settings. Below the header, there are three main panels: "Transcript", "Objectives", and "Directions".

**Transcript:**

- Hi. Welcome to the Metropolitan train station. Where are you traveling to?
- I'm going to New York.
- Great! I can help you plan your trip to New York. What day would you like to depart?
- I want to leave at first my.
- [Typing]
- I have a couple of trains that day. Does the train leaving at 2 PM work for you?

**Objectives:**

- Ask how long the trip is.
- Tell her you don't want the ticket.
- Tell her you want the ticket.
- Ask if there is another train.
- Ask when the train will arrive.
- Ask her to repeat what she said.

**Directions:**

The central part of the interface shows a 3D-rendered female character in a blue jacket standing behind a desk. In the foreground, there is a notepad titled "Trip to NYC" with the following text:

- o May 1 - May 7
- o \$300 Budget

Next to the notepad is a train ticket for "HAMLET New York City" on "May 1, 8PM" in "Row V, Seat 139".

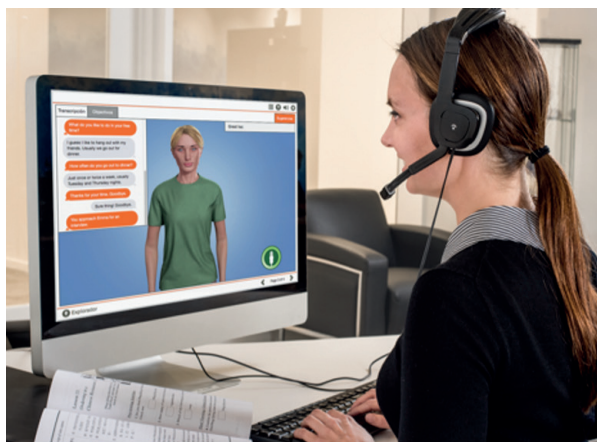
At the bottom of the interface, there is a navigation bar with "Explorer" on the left, "Developed in Partnership with LAUREATE EDUCATIONAL INC." in the center, and "Page 5 of 6" on the right with navigation arrows.

*Note.* Supportive features such as transcripts, learning objectives, and directions are displayed. @Alelo Inc. Used with permission.

The ECAs are represented as virtual humans, animated female or male characters that engage in contextualised dialogues with implicit constraints on meaning and form. Enskill provides various supportive features, such as scaffolding activities in simultaneous transcripts of the dialogue, directions for answer alternatives, and displaying the learning objectives (see Figures 7.5 and 7.6). This results in freedom for the student to either produce her/his own spontaneous utterances within the given context or use one of the alternatives suggested by the system. However, the pace of the virtual human's speaking or English variety is not changeable within the system. The theoretical foundation for the system is claimed to be influenced by instructional design, stressing that complex cognitive skills are best learned through a combination of whole-task practice and part-task practice (Johnson, 2019a). It is designed to be student-centred through opportunities for active speaking in learning activities, with the option of supportive features as scaffolding.

### Figure 7.6

*A student interacting with Emma, a female ECA in Enskill. A synchronous transcript and directions were offered a supportive features (scaffolding)*

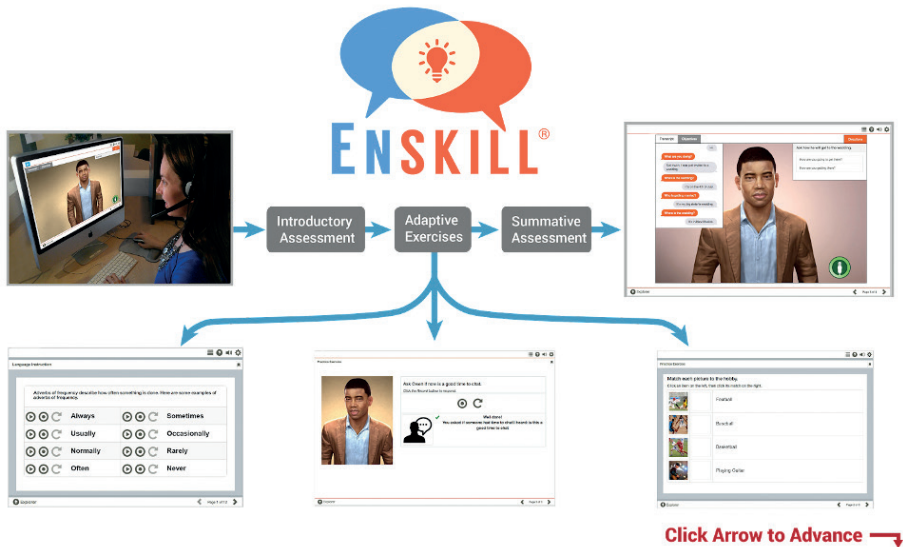


*Note.* @Alelo Inc. Used with permission.

The students' speaking skills are registered and analysed in the system by applying AI educational solutions (Johnson, 2019a). Auto-generated, individualised feedback on each student's speaking skills and adapted learning according to individual needs are part of the formative assessment learning process in the follow-up exercises, consisting of a variety of relevant grammatical content, vocabulary, pronunciation practice, and listening comprehension (see Figure 7.7). The SDS suggests different drill exercises, for example, drag-and-drop or pronunciation practice exercises, which are instantly corrected. The formative assessment of the system suggests that the student repeats the same spoken dialogue, exits, or moves on, depending on their success in the completion of the lesson. Additionally, the virtual human's social cues, such as body language, replies, or asking for the last utterance to be repeated, can be understood as feedback on the students' ability to make themselves understood in the target language in dialogue.

**Figure 7.7**

*Workflow of the students' activities in Enskill*



*Note.* An introductory assessment, supportive features, adaptive exercises, and summative assessment represent the workflow in Enskill (Alelo, 2022).

@Alelo. Inc. Used with permission.

A1- and A2-level conversational simulations include ten different themes; for instance, planning a party (A1) or buying a train ticket (A2). To pass the simulation, all the learning objectives must be accomplished. This SDS provides the students with an individual player’s dashboard (see Figure 7.8), showing individual information such as objectives finished/started, time, mastery score, and turns per minute).

**Figure 7.8**

*Screenshot of the view of the player’s dashboard in Enskill*

Date	Objectives <sup>1</sup>	Mastery Score <sup>1</sup>	Turns Per Minute <sup>1</sup>	Time Spent <sup>1</sup>
January 8, 2020, 10:18	0/3	0%	5.22	01:32
January 8, 2020, 10:11	3/3	73%	4.56	01:19
January 8, 2020, 10:08	3/3	74%	4.62	01:44
December 16, 2019, 12:10	3/3	47%	2.94	03:24
December 5, 2019, 10:01	3/3	69%	4.35	02:18

*Note.* @Alelo. Inc. Used with permission.

Mastery score is an automatically calculated measurement based on a combination of accuracy, fluency, and turns per minute, all based on the student’s performance in the simulation.

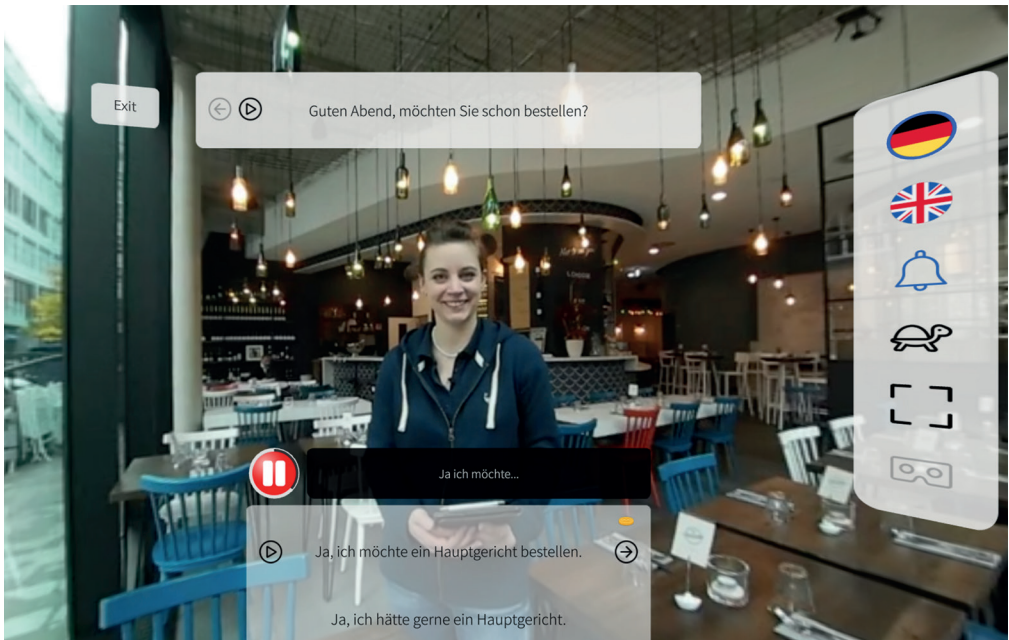
#### 7.1.4.2 IMMERSEME

ImmerseMe, used in Study 3, provides everyday-life scenarios in a VR learning environment in combination with Google ASR (ImmerseMe, n.d.). The students practised L2 speaking (German) with a pre-filmed native speaker as ECA in 360-degree videos. The ECAs were actual recordings of speakers’ “culturally authentic contexts” (Berti, 2020, p. 327), such as a restaurant (see Figure 7.9), a bakery, and at the zoo, where students could engage from various angles, by dragging the mouse around. Head-

mounted devices can be added to the speaking practice, such as Oculus Go, but were not applied in the study. ImmerseMe was initiated by a startup company in New Zealand.

**Figure 7.9**

*Screenshot of student view: Practising speaking German with a pre-filmed native speaker as ECA in a 360-degree video environment in ImmerseMe*



*Note.* The transcript of the spoken interaction is displayed, as well as supportive features such as language translation, sound effects, mode, and full-screen options. @ImmerseMe. Used with permission.

As a first step in ImmerseMe, the student chooses one of three levels (beginner, intermediate, or advanced), the theme and dialogue to practice, and in which mode. There is one choice, or sometimes an option of several utterances with distinct meanings, to say out loud in the spoken dialogue.

Hence, a more framed dialogue, with explicit constraints on meaning and pre-set constraints on the form. The next turn of the dialogue is already set, the student has to answer with exactly one of the alternatives displayed on the screen. The SDS offers some supportive features, such as the possibility of regulating the speed of the ECA's utterances. There are four learning modes for practising the speaking exercises, scaffolded with various support.

In the first mode, (i) pronunciation, the student answers the ECA's questions by saying exactly the suggested answer correctly so the ECA understands. Mistakes are shown in the speech-to-text field at the bottom of the screen. In the second mode, (ii) typing (dictation), the student listens to the ECA and then writes down what they heard, before choosing the correct utterance to answer with. In the third mode, (iii) translation, the answer is given in English, and the students translate it into the target language. The final mode, (iv) immersion, is where the learner has to apply everything that he or she has practised in the earlier modes. Ultimately, the goal is responding in the target language, knowing the answer by heart, and pronouncing it correctly to make it understandable to the ECA (as per the script). There is also still a scaffolding possibility to see a short English explanation. The system also offers an option in each situation for the student to click on key grammar and vocabulary associated with the chosen situation.

The Audio-Lingual method (Richards & Rodgers, 2001) of listening and repeating is one of the main theoretical ideas within this system, together with inspiration from Conti and Smith (2016), and refined by Conti (2022). Their stages of language learning were summarised into the system as: (i) listening as modelling (to raise awareness), (ii) thorough processing, (iii) structured production, and (iv) autonomy and spontaneity (ImmerseMe, n.d.). The division into learning modes for progressive learning and scaffolding skills development is understood as building on constructivist ideas (Huang et al., 2019; Meri-Yilan, 2019), preparing the student for full autonomy and spontaneous L2 speaking outside an educational context. After the student's successful spoken utterances, there are various positive reinforcements given, such as sound effects and words in bright colours flying around on the screen (e.g. KABOOM and WOW!).



On the main page of the system, the student can follow the total score he or she has gained after successfully completing the tasks. The generated information consists of time, activity, performance, and accuracy, which are displayed for the student's teacher in a teacher dashboard, as shown in Figure 7.10. Accuracy refers to the number of correct responses/total responses, expressed as a percentage.

**Figure 7.10**

*Teacher's dashboard displaying students' results and logged time in ImmerseMe*

2018-19 Research Participants - Top 50 Students

**Languages:** All Languages (12) ▾    
**Modes:** All Modes (13) ▾    
**Levels:** All Levels (10) ▾    
**Start Date:** Select a start date... ▾    
**End Date:** Select an end date... ▾

All time     Games Time

Reset Update Export All Values

Show  entries     Search:

Name	Total Points	No. of Responses	Response Accuracy	Active Time	Total Time	Lesson Modes Started	Lesson Modes Completed	Number of Logins	Last Log
S2 Participant, Student 22	216	357	77%	59m	-	42	26	1	08 Thu 2020 C
S2 Participant, Student 26	0	0	-	-	-	0	0	0	-
S3 Participant, Student 3	52	144	63%	24m	-	14	8	1	08 Thu 2020 C
S1 Participant, Student 1	125	243	76%	40m	-	26	18	1	08 Thu 2020 C
S1 Participant, Student 10	174	272	73%	45m	164d 57m	36	26	1	08 Thu 2020 C
S1 Participant, Student 11	72	156	83%	26m	-	19	14	1	08 Thu 2020 C
S1 Participant, Student 12	34	101	35%	16m	-	20	5	1	08 Thu 2020 C
S1 Participant, Student 13	85	146	68%	24m	-	24	14	1	08 Thu 2020 C
S1 Participant, Student 14	282	390	87%	1h 5m	-	48	39	1	08 Thu 2020 C

*Note.* ImmerseMe (n.d.). @ImmerseMe. Used with Permission.

In terms of the interaction, there are strong constraints on both meaning and form. In three of the four learning modes, the student must verbally repeat exactly what is suggested and expected by the system. Otherwise, the system does not understand, there is a communication breakdown, and the conversation stops abruptly. The system will continue the dialogue if the student has finished the suggested utterance adequately or chooses to skip that part and move to the next turn.

#### 7.1.4.3 SIMILARITIES AND DIFFERENCES BETWEEN THE TWO SDSs

The functionalities of the two selected SDSs that have a bearing on the results and analyses of Studies 2–4 of this thesis are highlighted in this subsection (see Table 7.1). The comparison of the two systems builds on evaluation criteria adapted from CALL frameworks (Bibauw et al., 2019; Rosell-Aguilar, 2017; Sofkova Hashemi et al., 2017), whereof Bibauw et al. (2019) have provided the majority of categories used. The categories that related to more of a subjective user experience were evaluated by the students in the current studies. Both met the thesis’ set-up criteria for the selected SDSs in Studies 2–4.

In both systems, the dialogues were “system-initiated”, meaning that they used “directive prompts” (Bibauw et al., 2019, p. 21) or asked leading questions to initiate the conversation. In Enskill, the interactional implication of initiative management was initiated by the system, allowing mixed initiative from the user, whereas in ImmerseMe, the system led the conversation using directive prompts. Both systems were interactive, applying some kind of ECA to simulate a human interlocutor. Enskill (used in Studies 2 and 4) uses intelligent animated virtual humans as ECAs, whereas ImmerseMe (used in Study 3) uses pre-filmed human native speakers as ECAs, enhanced by low-immersive technology, VR.

Both systems were “goal-oriented” (Bibauw et al., 2019, p. 22) towards solving a task in everyday-life situations where the target language was spoken, and links can be drawn to TBLT (see Section 2.1.2). The systems situated the language learning in everyday-life situations using technology intended to immerse the student and provide them with the experience of being in that location. Both systems used multimodal interfaces in various ways, such as sound and visual effects.

They provided possibilities for spoken conversation practice and some kind of text-to-speech while simultaneously displaying a written transcript of the spoken utterances. In Enskill, there was an option to display a transcript of the ongoing spoken dialogue. In ImmerseMe, the student's spoken utterance was transcribed and displayed simultaneously depending on selected mode. The spoken interaction was instantly followed by implicit feedback. If the dialogue was ongoing, it meant that the CA had understood the student, which indicated that his/her speaking skills were accurate according to what the system could expect in that context. Neither of the systems gave explicit feedback on pronunciation or offered a dialect variety. There was little focus on learning about other cultures in either system. However, in ImmerseMe, some scenarios were filmed in areas where the target language to be practised is spoken.

The major difference between the systems that is important for the L2 speaking practice lies in the constraints. These can be on meaning, limiting the student in terms of the content and/or form they can express, and hence limiting their use of the language (grammar and vocabulary). This has pedagogical and interactional implications. Constraints follow a continuum of four levels: pre-set form/meaning, explicit, implicit, and no constraints (Bibauw et al., 2019). Enskill provides the student with possibilities for the spontaneous production of utterances, with implicit form and meaning constraints in accordance with the previous turn in the spoken interaction with the CA. In contrast, ImmerseMe, which places explicit constraints on form and meaning, offers students the possibility of practising reading aloud one utterance, chosen as either the only given alternative or as one among a few utterances on a list, or writing it, depending on the mode selected. The system has a predetermined script, and the student repeats verbatim prompts only in educational contexts. The level of constraints regulates the complexity of the student's own production and construction of utterances.

**Table 7.1**

*Overview of similarities and differences between Enskill and ImmerseMe*

<b>Design</b>	<b>Enskill</b>	<b>ImmerseMe</b>
<b>Interactional Criteria</b>		
Constraints on Meaning	Implicit Implied naturally from the dialogue situation, the previous turn of the speaker, and the task to be solved. ECAs do not retain the meaning of utterances spoken during earlier turns in the dialogue.	Explicit
	Spontaneous speech within a context (scenario)	List of utterances (prompts) to choose from
	Task to be completed Goal-oriented	Task to be completed Goal-oriented
Constraints on form	Within context	Explicit List of utterances (to be read out loud)
<b>Technological Aspects</b>		
Meaning processing	NLP analyses students' output and generates appropriate ECA responses in the next turn. ASR adapts the ECA's utterances in relation to the student's previous utterances.  Text-to-speech	Validating expected meaning according to a predetermined script with fixed utterances. ASR recognises the student's pre-set utterance or their choice from two or many, and generates the fixed ECA utterance in the next turn.  Text-to-speech
Data-driven vs rule-driven	Data-driven, iteratively developing in accordance with the spoken interaction between the system and the users, and probabilistic.	Rule-driven (handcrafted)

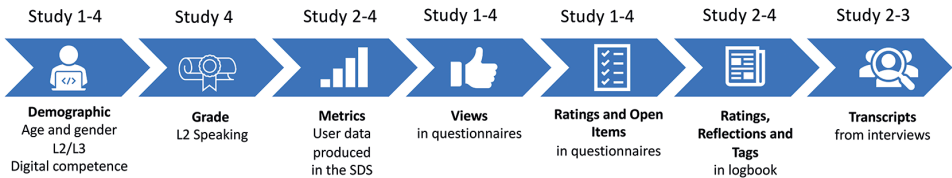
The SDSs guided and facilitated the learning activity, including the ECA as the interlocutor speaking the target language and the student him-/herself self-regulating through the choices of modes and levels, and independently choosing between the supportive features and activities provided in the system. The metrics of the two SDSs provided quantitative data about the students' practising of speaking skills and performances (for instance, time, trials, fulfilled objectives, and mastery score/points), i.e. summative feedback.

## 7.2 DATA PRODUCTION

A combination of data production methods was selected to produce different data types in Studies 1–4 (see Figure 7.11) and gain information to answer RQs 1–3. This section outlines the methods and gives a more detailed picture of the data production.

**Figure 7.11**

*Overview of types of data produced in the studies*



Methods that produce these forms of data about the students' experiences of practising speaking skills in SDSs in Studies 2–4 were chosen to align well with the students' educational practice and curricula. Using a variety of methods offers a comprehensive approach, enabling qualitative data to extend and enrich quantitative data, and vice versa, to enable a deeper understanding of the object of study (Bergman, 2008; Creswell & Creswell, 2018; Levy, 2015). Metrics generated in the SDSs, along with ratings in the questionnaires (Johnson, 2019a), were used as quantitative data. Then, in order to nuance and deepen these insights open items in the questionnaires, written logbook reflections, and interviews were used to

clarify the quantitative results' underlying aspects of importance. By incorporating diverse methods, valuable insights can be uncovered, ultimately enriching the validity and richness of the thesis' conclusions. The following selected methods are described and motivated in this section:

- Metrics reports generated in the SDSs
- Teacher questionnaire
- Student questionnaires
  - pre- and post-trial (experiences of practising speaking in an SDS)
  - initial- and final (experiences of speaking with ECAs)
- Students reflecting in digital logbooks (Google Docs & LoopMe)
- Teacher and student semi-structured interviews

### 7.2.1 METRICS

Metrics reports from the students' speaking activities and results in the SDSs were collected to provide quantitative data about their performance. Logged data sets in the SDSs consisted of the students' login times, trials, lessons completed, learning objectives fulfilled, response accuracy, and mastery score or points. These metrics were exported and organised in Excel for further analysis in SPSS 28.

### 7.2.2 TEACHER QUESTIONNAIRE

A questionnaire with 16 items revolving around L2 teachers' views and experiences of digital tools in L2 education (RQ1) and demographic data was distributed to the participating L2 teachers in Study 1. A few items had sub-questions and optional free-text comments. Digital competence was included in terms of the teachers' ability to support students in their use of digital tools. A Likert scale was used, with ratings of 1 to 5, where 1 = strongly disagree and 5 = strongly agree or, in some contexts, rating 1 = very seldom and 5 = very often, or 1 = not at all and 5 = to a great extent. multiple-choice items, where one or many answers were possible. In line with Brown (2009) and Dörnyei and Taguchi (2010), an open-ended question with the possibility for free comments was provided. This

questionnaire was tested in the Pilot and refined before Study 1. As suggested by Creswell and Creswell (2018) as well as Dörnyei and Taguchi (2010), the use of a questionnaire as a method provides an efficient way of exploring participating teachers' views and experiences simultaneously and within a limited time.

### 7.2.3 STUDENT QUESTIONNAIRES

Two digital questionnaires, pre- and post-trial, operationalising concepts from RQs 2 and 3 about how students experience practising L2 speaking skills in an SDS and speaking with ECAs were iteratively developed into four questionnaires (Q1–Q4). These contained fine-grained and systematic items in line with the FoSCAI and the dimensionalising of experiences. Speaking skills were sub-categorised into pronunciation of phrases, one's own production of utterances, and interaction (including dialogue). The formulations in the questionnaires were inspired by the questionnaires in studies by Johnson (2019a) and Erickson et al. (2022). The categories and indicators of the RoI questionnaire (Stenbom, 2015, p. 34) provided inspiration for the design and the items of Q2–Q4.

The choice of a questionnaire is useful when the focus is on participants' experiences of the object of study (Creswell & Creswell, 2018; Dörnyei & Taguchi, 2010), and here it was also used in combination with other instruments. Applying the same items pre- and post-trial facilitates the investigation of the object of study over time. Items were prepared in language adapted to young people, as suggested by Källström and Andersson Bruck (2017). When applicable, illustrations or screenshots from the system were used to help the students understand the terms used in the various items, e.g., transcripts in the SDS. Before the questionnaires were distributed in the studies, the instrument was tested on students of the same age as the participating students and a few items were refined regarding choice of words. All the student questionnaires are presented and motivated next.

### 7.2.3.1 PRE-TRIAL QUESTIONNAIRE (Q1)

The pre-trial questionnaire was conducted in order to gain an understanding of the students' (i) views on L2 learning, especially speaking skills, (ii) views on L2 speaking and confidence, and (iii) experiences of using digital tools. It started with a few questions about demographic information, such as gender, age, and questions about the daily use of digital devices.

In the first version of the pre-trial questionnaire, there were 30 background questions relating to the target language (English or German, respectively). The items were based on SLA principles and ideas that are considered important for effective L2 learning and development (see Section 4.4). In all versions, a Likert scale was used, ranging from 1 to 5, where 1 = strongly disagree, and 5 = strongly agree. In some items, rating 1 = very seldom, and 5 = very often, or 1 = not at all and 5 = to a great extent, depending on the question. Some items had a subsection providing the opportunity to write free comments, or several given options where one or many answers were possible multiple-choice items. There were also open-ended items with free comments.

### 7.2.3.2 POST-TRIAL QUESTIONNAIRE (Q4, REFINED VERSION)

In this thesis, there were two versions of the post-trial questionnaire, which were designed before and after FoSCAI. The first version of the post-trial questionnaire (used in Studies 2 and 3) consisted of 18 questions about the students' experiences of practising English speaking skills in the SDSs and their self-assessment of learning and development of speaking skills. The questions were based on the SLA principles that are considered important for L2 learning and development (see Section 4.4) and the outcome was a follow-up from the pre-trial items. Pictures were used in the questionnaire to highlight the difference between views of learning to speak English and speaking English, respectively, and of educational digital tools. The post-trial questionnaire was conducted in order to gain an understanding of how the students practised in the SDS based on their self-reported experiences of using the selected SDS. The Likert scale was used in the same way as in the pre-trial questionnaire.

In the refined version of the post-trial questionnaire (Q4), developed in line with FoSCAI, there was a further development of the items, making



up a total of 45. There were new specified items representing the teaching dimension which asked about the students' experiences of using the supportive features (i.e. transcripts, objectives, directions, practise exercises, feedback, and the learner's dashboard). The RoI questionnaire, which is used for human-to-human online learning (Stenbom et al., 2016), was used as inspiration for the formulation of items within the four dimensions exploring the students' experiences of practising L2 speaking in an SDS (RQ2) and speaking experience with ECAs (RQ3). This was separately explored in the initial and final questionnaire (Q2 and Q3), as described next.

### 7.2.3.3 INITIAL AND FINAL QUESTIONNAIRE ON SPEAKING WITH AN ECA (Q2 AND Q3)

These were two identical questionnaires, addressing initial and final experiences of speaking with ECAs (Q2 and Q3), with just the tense of the verbs changed from present to past tense (e.g., How was it to SPEAK English with an avatar? capitalised originally). The items focused on the students' experiences dimensionalised into cognitive, social, and emotional aspects. The battery of items was derived from the RoI questionnaire about one-to-one online learning (Stenbom, 2015), with the statements adapted to this educational context. Items measuring student experiences of conversational interfaces, here operationalised through ECAs, were redesigned based on terms drawn from user experiences of conversational interfaces (Kocaballi al., 2018).

Since the experiences of speaking with ECAs were considered a central part of the students' practice in an SDS for this thesis, they were explored separately in Q2 and Q3 in order to gain a nuanced understanding over time, and without becoming mixed together with a focus on L2 learning and development as in Q4. The earlier questionnaire (Q2) was about the students' initial speaking experiences with ECAs. It included ratings with open-ended questions and questions offering multiple-choice answers in a total of 23 items. The final questionnaire (Q3) was about students' accumulated speaking experiences with ECAs, and included the same items as Q2. Both Q2 and Q3 used the same Likert scale as the previously described questionnaires (Q1 and Q4).

## 7.2.4 DIGITAL LOGBOOKS – SHARED DOCUMENT AND LOOPME

Data was also generated through two types of digital logbooks where students recorded written self-reports about their experiences. These were filled out immediately after the practice in the SDS. The idea is similar to an exit ticket (William, 2017) with reflective questions about L2 development and learning, and the materials and methods used during the speaking session. The students reflected for about five minutes.

### 7.2.4.1 SHARED DOCUMENT

An individual open shared Google document was used as a digital logbook for the students to engage in free reflection about their speaking session. The content of these reflections inspired some of the questions included in the semi-structured interviews, allowing the exploration of some issues in even greater detail, providing that they were in line with the overarching RQs.

### 7.2.4.2 LOOPME

The digital logbook was developed in the system LoopMe (Ericsson, 2021; Lackéus, 2020; LoopMe, 2023), as visualised in Figure 7.12. The students produced data in three ways, in line with the idea of combining multiple methods. Firstly, after each speaking session, the students answered questions. Secondly, they rated their overall experience of the speaking session with an emoji of a smiling or frowning face corresponding to a five-point Likert scale from very bad (-2) to very good (2), where 0 is the neutral mid-point. Thirdly, the students chose from a battery of 27 suggested tags (representing feelings or goals to reach) corresponding to their SDS experiences.

Data on the students' self-reported experiences was produced systematically and quantitatively, already coded by the students themselves through free-text tagging and not by the researcher. The tags were selected and prepared based on an amalgam of the aims of L2 speaking in Swedish lower-secondary schools and key SLA principles for L2 learning (see Chapter 4) to address RQs 2 and 3 about students' experiences of SDS practice and their L2 speaking experience. FoSCAI guided the preparation of the tags and reflective questions. These questions and tags were written in language that was adapted for the students and their understanding of the content.

**Figure. 7.12**

*Example of a reflective post in the digital logbook LoopMe (my translation from Swedish)*

**Reflection after session 1**

**A. Answer the following questions:**

1. How was it to speak English with an avatar today? Develop your answer, why do you think it was like that?
2. What have you worked with in today's tasks in Enskill? For instance situation(s), content, supportive features (transcripts, objectives, direction) you have used, which follow up exercises you have done, did you take part of your results in the Player Dashboard?
3. What have worked well during your speaking session?
4. What have worked less well during your speaking session?

**B. Feelings**

Click on an emoji that corresponds to your overall feeling of the experience of the task/s/ you have worked with today in Enskill.

**C. Tags**

Mark suitable "tags", words describing how you experienced practising in Enskill, what you think you have practised and the result of it. Mark as many tags as you feel correspond to your experience today. You have to choose at least one A-tag and at least one B-tag.

A. Engaged/ motivated	A. Frustrated	A. Felt social connection w.avatar	
A. Easy	A. Meaningful	A. Meaningless	
A. Dissatisfied	A. Nervous/ worried	A. Disengaged/ unmotivated	
A. Fun	A. Conversation w. avatar felt natural	A. Exciting	A. Safe
A. Stressful to have enough time	A. Difficult	A. Boring	
A. Challenged on my level	A. Willing to to speak	A. Satisfied	
B. Self-regulated how I practised (speaking)	B. Answered questions	B. Listened to understand	
B. Fulfilled all learning objectives in the task	B. Spoken in everyday life scenarios		
B. Asked questions	B. Practised my pronunciation	B. Practised conversation/dialogue	

*Note.* The reflection post consisted of: (i) open-ended items, (ii) emojis to rank from -2 to +2, and (iii) tags to choose from representing feelings or goals to reach corresponding to individual experiences of the SDS. Adapted from LoopMe (2023)

LoopMe has previously been applied within the research and practice of entrepreneurship and enterprise education (Lackéus, 2020) and among teachers, and is further described in Lackéus' handbook for research in schools and preschools (2021). It is "optimized for social science purposes and used primarily for data collection and analysis" (Lackéus, 2020, p. 200) and is based on a development of the Experience Sampling Method, which is a research method for producing longitudinal data through the repeated sampling of various aspects of people's daily experiences and behaviours in natural everyday-life settings (Hektner et al., 2007). Participants are prompted to report on their current thoughts, feelings, and activities, providing a detailed and valid understanding of their experiences as they occur in daily life. Experience Sampling Method allows researchers to capture fluctuations, assess within-person variability, and examine the contextual factors that influence individuals' subjective experiences.

#### 7.2.5 SEMI-STRUCTURED INTERVIEWS

Semi-structured post-trial interviews were conducted with the students in pairs. The selection was a convenience sample based on those students who consented to participate in a recorded interview, and the total number of interviews was decided iteratively, once data saturation had been reached through redundancy signalling. Four thematic pictures were used to guide the interviews. Details from the students' answers were sometimes repeated in follow-up questions to enable further understanding (Källström & Andersson Bruck, 2017). The recorded interviews were transcribed, anonymised, and then coded manually. The researcher (myself) also took notes during the interviews. The analysis of transcript data from the interviews enabled a detailed understanding (Creswell & Creswell, 2018) of how some students experienced practising speaking skills in an SDS, described in their own words. It gave the possibility to ask further questions and gain "access to the meanings people attribute to their experiences and social world" (Miller & Glassner, 2016, p. 53).

Semi-structured interviews were also conducted with the teacher (and, in one case, a trainee teacher), both before and after the trials. These interviews were conducted in order to understand, from the teacher's perspective, how the students practised their speaking skills with an ECA in an

SDS and used its supportive features, practical and technical issues that arose during the trial, and what to improve for the subsequent study. All the recorded interviews were transcribed, anonymised, and coded manually. The researcher (myself) took notes during the interviews. The data was used mainly for input into the redesign of Study 4 and analysis of data from Study 2, which is reported in Paper 2.

### 7.3 METHODS AND DATA IN STUDIES 1–4

This section first provides an overview of the Pilot and included Studies 1–4 (see Table 7.2) described in relation to the corresponding RQ(s) of the thesis, and the aim of the paper, methods and data, participants, and the study's outcome. Secondly, each study is delineated in detail, with a description of its design, realisation, and methods of data production. Then follows a summary of the lessons learnt and a section highlighting the role of the teacher and the researcher in this thesis. As a reminder, the aim of this thesis is to: Explore and gain a nuanced, critical, and comprehensive understanding of how conversational AI is experienced by students during their institutional language education when practising speaking skills through face-to-face interaction. The three overarching RQs are:

RQ 1: Focusing on speaking skills, what are language teachers' views and experiences of teaching and using digital tools?

RQ 2: How do lower-secondary students experience practising L2 speaking skills in an SDS?

RQ 3: How do lower-secondary students experience speaking L2 with an ECA?

**Table 7.2**

*Overview of Studies 1-4*

	Studies	Aim + Overarching RQs	Methods and data	Participants	Outcome
TEACHERS	<b>Pilot:</b> Spanish Teacher Meeting January, 2019	Explore the Swedish L2 educational situation, focusing on speaking skills, teachers' views and experiences of teaching and using digital tools <b>RQ1</b>	-Questionnaire -60 min audio-recorded workshop  Ratings and open-item answers ( $N = 9$ ) Transcription (representative parts) and notes from audio-recordings of workshop	16 teachers	Tested questionnaire Test of Enskill and ImmerseMe  Input for Study 1
	<b>Study 1:</b> L2/L3 Language Teacher Meeting March, 2019	Same as in the Pilot  <b>RQ1</b>	Questionnaire  Ratings and open-item answers ( $N = 36$ )	36 teachers	Paper I
STUDENTS	<b>Study 2:</b> <b>iSpeak_Eng</b> Students speaking English with virtual humans in Enskill May 10-28, 2019	Explore the perspectives of lower-secondary students on practising their English language speaking skills through verbal interactions with virtual humans in an SDS (Enskill).  <b>RQ 2 and RQ3</b>	-Pre- and post-trial questionnaires -Semi-structured interviews -Logbooks -Metrics  -Open-item answers and ratings ( $N = 25 + N = 23$ ) -Students' written reflections ( $N = 22$ ) -Teacher's written reflections ( $N = 1$ ) -Recordings and transcriptions of interviews with students in pairs ( $N = 6$ ) + the teacher ( $N = 1$ ) and trainee ( $N = 1$ ) -Metrics generated in Enskill -Transcripts (exemplifying quotes)	25 students	Paper II
	<b>Study 3:</b> <b>iSpeak_Ge</b> Students practising speaking German with pre-filmed native speakers in ImmerseMe May 10-28, 2019	Explore experiences of Swedish lower-secondary students in a formal educational context practising German speaking skills with CAs in the desktop version of ImmerseMe.  <b>RQ2 and RQ3</b>	-Pre- and post-trial questionnaires -Semi-structured student interviews -Logbooks (students and teacher)  -Open-item answers and ratings ( $N = 41 + N = 39$ ) -Students' written reflections ( $N = 41$ ) -Teacher's written reflections ( $N = 1$ ) -Recordings and transcriptions of interviews ( $N = 4$ ) + teacher ( $N = 1$ )	41 students	Paper III
	<b>Study 4 *</b> <b>iSpeakL_Eng</b> (initially) Students practising speaking English with virtual humans in Enskill October, 2021	Expand the understanding of how lower-secondary students experience the use of conversational AI for spoken interaction.  <b>RQ2</b>	-Pre-trial questionnaire (Q1) -Initial questionnaire (Q2) -Logbook LoopMe -Metrics  -Open-item answers and ratings from Q1 ( $N = 22$ ), and Q2 ( $N = 21$ ) -Students' written reflections and ratings ( $N = 21$ ) and 2 ( $N = 22$ ) -Transcripts (exemplifying quotes)	(i) Initially ( $N = 22$ )	Paper IV
	<b>Study 4 *</b> <b>iSpeakL_Eng</b> (longitudinally) Students practising speaking English with virtual humans Enskill October 1, 2020 -February 12, 2021	Explore and gain a nuanced and comprehensive understanding of how conversational AI in an institutional L2 educational situation is experienced for the learning and development of speaking skills over time.  <b>RQ2 and RQ3</b>	-Pre-trial questionnaire (Q1) -Post-trial questionnaire (Q4). -Initial (Q2) and final (Q3) questionnaires -Logbook LoopMe  -Open-item answers and ratings (Q1-Q4). -Completed questionnaires ( $N = 20$ ) (Q1, $N = 22$ ; Q2, $N = 21$ ; Q3, $N = 21$ ; Q4, $N = 20$ ) -Students' written reflections and ratings -Transcripts (exemplifying quotes) -Grade (A-F)	(ii) Longitudinal ( $N = 21$ )	Paper V

*Note.* \* Study 4 was split into two parts, one initial part based on one early dataset and one longitudinal part based on a dataset produced longitudinally.

### 7.3.1 PILOT STUDY

In line with Ejlertsson's (2019) motivation of conducting a pilot study in order to test and refine research design and instruments for future studies, this Pilot study was conducted in preparation for Study 1. By engaging a convenience sample of 16 Spanish-language teachers during The Spanish Teacher Association's annual meeting in the West of Sweden in January 2018, completed questionnaires were collected for nine teachers. Missing data ( $N = 7$ ) can be explained by the fact that some teachers left the meeting before the questionnaire was distributed, or were retired and not able to answer questions about today's educational situation. During the meeting, there was a mixture of lectures, a demonstration and trial of two SDSs (Enskill and ImmerseMe) for practising L2 speaking with an ECA, followed by a discussion. Subsequently, the teachers answered the teacher questionnaire, which was also evaluated and refined before application in Study 1. Their answers were later analysed using descriptive statistics in relation to rankings, answers to open-ended items, and free comments. The workshop was audio-recorded for practical documentation. A summary of the findings was written to support the further direction of the studies that constitute this thesis and was also used as a validity and reliability control for the results of Study 1.

### 7.3.2 STUDY 1

Study 1 was conducted at an annual meeting of The Language Teachers' Association in the West of Sweden in the spring of 2019, and involved a convenience sample of 41 language teachers. They were between 32–69 years old and L2 (English) and L3 (Spanish, French, and German) teachers. Questionnaires were completed ( $N = 36$ ). Missing data ( $N = 5$ ) are explained by the fact that some teachers left the meeting before the questionnaire was distributed, or were retired teachers who were unable to answer questions about today's educational situation. The theme for the meeting was oral proficiency, framed by a lecture on the subject. Hence, the content of this study fitted well with that established context. The teacher questionnaire was distributed digitally but was also available in analogue form to the participating L2 teachers. The quantitative and qualitative data generated from their ratings and answers to open-ended items on the questionnaire relating to RQ1 were analysed using descriptive statistics. The results were reported in Paper I.

### 7.3.3 STUDY 2

The convenience sample consisted of Swedish secondary school students ( $N = 25$ ), aged 14–15 years, who were studying English. They practised speaking English individually in the SDS Enskill, in a classroom setting over a period of two weeks. The produced data consists of quantitative user data in metrics reports generated in the SDS and quantitative and qualitative data gathered through pre- and post-trial questionnaires and digital logbooks written up in a shared document. Semi-structured interviews were carried out and recorded with students who agreed to participate ( $N = 6$ ) in pairs at the end of the trial period. Interviews were also conducted with the teacher and the trainee teacher, both before and after the trial. The data was analysed using descriptive statistics and reflexive thematic analysis, as described in Section 7.4.2. More detail is provided in Paper II.

### 7.3.4 STUDY 3

The convenience sample consisted of Swedish lower-secondary students ( $N = 41$ ) drawn from two groups studying German, aged 13–15 years. They practised speaking German with pre-filmed human beings as ECAs in the desktop version of the SDS ImmerseMe, which employs 360-degree videos. The speaking sessions were spread out over a total period of three weeks. The produced data consisted of quantitative user data generated in the system and qualitative data from digital pre- and post-trial questionnaires and digital logbooks written up in shared documents. The design of the questionnaires and logbook was the same as in study 2, just replacing the word “English” with “German” in the questions. Semi-structured interviews with volunteer students ( $N = 4$ ) were conducted at the end of the trial period, as well as with the teacher, both pre- and post-trial. This study was reported on in Paper III.



### 7.3.5 REFINEMENT BEFORE STUDY 4

Based on lessons learnt from Studies 2 and 3, there was an identified need to systematise further and operationalise the students' experiences. This is the reason why the FoSCAI was developed. Theoretical, empirical, and learning-situation-based arguments were found in common with RoI, which inspired the design of FoSCAI. Consequently, RQs 2 and 3, the instruments, the organisation of data production, and the analysis were adjusted and refined. The students' experiences of practising speaking skills (RQ2) in an SDS were dimensionalised, and the experience of the central speaking activity with ECA was extracted (RQ3) for all the students together, giving their educational experience of practising L2 speaking in an SDS.

The decision to use Enskill again was based on experiences from Study 2 at both a technical and a practical level. Consequently, Study 4 was conducted longitudinally and further integrated into the educational framing. The simulations were chosen pre-trial by the teacher (in dialogue with the researcher) and organised into themes for integration into the lesson plan. The students were given refined instructions and told to practise the simulated situations as many times as necessary until the learning objectives were fulfilled. They were also encouraged to explore the supportive features of Enskill, which scaffold the student through the speaking activity by means of direct feedback and formative assessment with individualised practice exercises. Furthermore, in line with FoSCAI, the questionnaires (Q1–Q4) and the logbook were designed for refinement, providing a more systematic and approachable format for the participants to share their ratings and reflections over time. The two sets of questionnaires were developed in order to separate the experience of the central L2 speaking experience with an ECA for detailed analysis, regardless of the learning objectives. The students' educational experiences in the SDS were explored by means of their self-estimated overall experience (emoji in the logbook and in Q4), including the experience of speaking with an ECA (Q2 and Q3) and students' ratings of the teaching dimension post-trial (Q4).

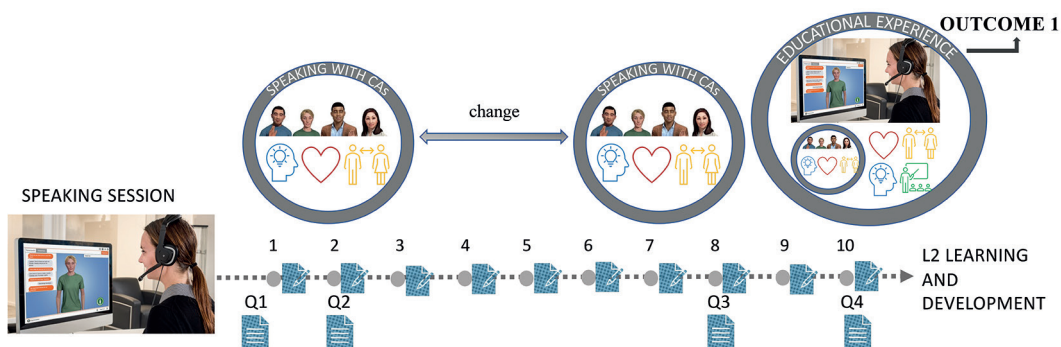
### 7.3.6 STUDY 4

The convenience sample consisted of Swedish lower-secondary students ( $N = 22$ ) studying English, aged 13–14 years, consisting of 13 male and nine female students. One of the students quit the class during the trial, giving  $N = 21$  for the longitudinal data. Three students in the class chose not to participate in the study at all. Swedish lower-secondary school students were referred to as Middle-school students, written to American standards according to the relevant journal's preferences. The students practised speaking English with embodied CAs in the SDS Enskill.

The speaking sessions were spread over a period of four months (including a break for Christmas), and revolved around three themes (music, travel, and food). As visualised in Figure 7.14, the longitudinal design of this study applying the FoSCAI included ten repeated speaking sessions over time in Enskill, which were immediately followed up with self-reports in the digital logbook LoopMe. These generated the outcome referred to as educational experience. Additionally, the students answered two sets of questionnaires, pre- and post-trial questionnaires, and initial and late speaking experience questionnaires (Q1, Q4 and Q2, Q3, respectively, see Section 7.2.3). The SDS's metrics were only used to track the students' speaking activities, such as time spent, and started and finished learning objectives. Background variables, such as gender and proficiency (based on the teacher's assessment of L2 speaking), were collected to compare results in subgroups.

**Figure 7. 14**

*Longitudinal design of Study 4 with ten speaking sessions generated the outcome educational experience*



*Note.* CAs = conversational agents (here animated virtual humans)

1-10 = speaking sessions and measure points



Q = questionnaire (Q1-Q4)



= digital logbook LoopMe

Data around the experience of speaking with CAs was produced and analysed separately for inclusion in the outcome and educational experience. The speaking activity is designed to lead to L2 learning and development.

The speaking sessions were used as a supplemental educational activity in English lessons to provide speaking practice in the target language, organised in themes. It was integrated into the ordinary lesson plan, inspired by Goh and Burn's teaching–speaking cycle (2012, p. 153). The sessions focused on speaking skills, the production of utterances and interactions in spontaneous speech, and repeated tasks, followed up by reflection (in log-books) and feedback (by the CA and SDS). Due to the COVID-19 pandemic, the study was carried through without the researcher being physically present in the classroom at any time, just online at the last time point. The pandemic forced an involuntary break, and the students' speaking sessions took place in diverse settings. The students were in school initially and finally, but also performed a few online speaking sessions at a distance. Due to the situation, the dataset was split into two sets. The initial dataset, after two speaking sessions, was analysed using descriptive statistics and reflexive thematic analysis, whereas the longitudinal data set was analysed using inferential statistics. The results were reported in Papers IV and V.

## 7.4. DATA ANALYSIS

Various data types were produced in the studies, from a combination of methods. In Table 7.3 there follows a description of the methods of analysis, study by study (for the results, see Chapter 8). Finally, this chapter discusses ethical concerns and the role of the teachers and the researcher in Studies 2–4, as well as the role of the researcher in Study 1.

**Table 7.3**

Data generated in relation to each study of this thesis and the corresponding methods of data production and analysis

Study	Participants	Analysis	Analysed data	Metrics	Questionnaire	Logbooks	Interview
Pilot (Study 1) Workshop	16 teachers	Descriptive statistics	-Ratings and open-ended items -60 min audio-recorded workshop	-	N = 9	-	-
Study 1 Meeting	36 teachers <sup>3</sup>	Descriptive statistics	Ratings and open-ended items	-	N = 36	-	-
Study 2 iSpeak_Eng	25 students (1 teacher + 1 teacher trainee)	Descriptive statistics Reflexive thematic analysis	-Metrics generated in Enskill -Ratings and open-ended items -Written reflections -Recordings and transcripts of interviews -students, total time 34 mins -teacher and trainee teacher, total time 71 mins	N = 25	Pre-trial N = 25 Post-trial N = 23	25 Students x4 posts/pers 1 Teacher x 4 posts/pers	6 students in pairs 1 teacher + teacher trainee
Study 3 iSpeak_Ge	41 students	Descriptive statistics Reflexive thematic analysis	-Metrics generated in ImmerseMe -Ratings and open-ended items -Written reflections -Recordings and transcripts of interviews -students, total time 21 mins -teacher, total time 71 mins	N = 41	Pre-trial N = 41 Post-trial N = 39	41 Students x 9 posts/pers 1 Teacher 4 posts	4 students <sup>2</sup> (pairs) 1 teacher
Study 4 iSpeak_Eng <sup>5</sup>	22 students <sup>4</sup>	Descriptive statistics (i, ii)	-Metrics generated in Enskill -Ratings and open-ended items -Written reflections, ratings, and tags (LoopMe) -Pre-trial grade (speaking) Inferential statistics (ii) FoSCAI (i, ii)	overall, total N = 21	1. Pre-trial (Q1) N = 22 2. Initial (Q2) N = 21 3. Final (Q3) N = 21 4. Post-trial (Q4) N = 20	Total 22 (i) x 2 posts/p (ii) x 10 posts/p	-
Total	52 teachers 88 students			87 pers	257 questionnaires	36 logbooks	10 persons

2 Whereof two students participated in both studies and interviews

3 One teacher was benevolent to participate in the student studies.

4 One student left the class during the study. Complete data produced for 20 students longitudinally.

5 Data were split into two subsets (i) initial data

#### 7.4.1 STUDY 1

In Study 1, the teachers' ratings in the questionnaire were analysed using descriptive statistics. The number of teachers and mean values for various items were calculated. The background variable of teaching language (English, French, German, and/or Spanish) was used for subgroup analysis to identify possible patterns.

#### 7.4.2 STUDIES 2–4

In Studies 2–4, the quantitative data gathered from the questionnaires and SDSs metrics were imported into Excel for organisation and then further exported into SPSS 28 (Field, 2018) for descriptive statistics to calculate the mean (*M*), standard deviation (*SD*), and median value (*Mdn*). Possible connections between variables in the various datasets were analysed. Correlation plots were conducted. In Studies 2 and 3, the analysis of the questionnaire data was done after the interviews in order to avoid the risk of bias from the given answers.

Studies 2 and 3, as well as the initial part of Study 4, added the use of reflective thematic analysis, in line with Braun and Clarke (2006, 2019; see further description in Papers II–IV). The analyses were conducted on data from the students' written logbook reflections, answers to the open-ended items from the post-trial questionnaire, and in Studies 2 and 3, also transcripts of student interviews, all of which were initially coded by the researcher with collegial assistance. In Study 3, the data was coded systematically as a preparation until a collegial consensus was reached with another researcher's separate coding. A combination of data-driven (inductive) and theory-driven (deductive) approaches were used (Braun & Clarke, 2012), and themes were developed in relation to the overarching RQs 2 and 3 of the thesis (see Chapter 8). During the analysis, these themes were refined and given illustrative titles (Nowell et al., 2017). As Brown (2009) suggested, a selection of representative quotes from the student reflections was used to further illustrate and add depth to the results presented in Papers II–V.

### 7.4.3 STUDY 4

Additionally, in Study 4, a methodological and analytical lens based on FoSCAI was employed and the focus was on the analysis of quantitative data (for the content of variables, see Table 7.4). Due to the COVID-19 pandemic, the data analysis was divided into two subsets: (i) initial experiences of L2 speaking with ECAs, and (ii) longitudinal educational experiences of practising L2 speaking in an SDS. The two data sets were connected to the same demographic data produced in the pre-trial questionnaire (Q1). The results were reported in Paper IV and Paper V, respectively (see Sections 8.5 and 8.6).

**Table 7.4**

*Overview of the content of variables based on items from the two sets of questionnaires (Q1–Q4 and Q2–Q3), and data from the logbook LoopMe*

Pre-trial Q1	Initial (ECA) Q2	Late (ECA) Q3	Post-trial Q4	Logbook
-Demography -Educational digital tools -Views on learning speaking and views on speaking English	(1) Our calculation of the speaking experience based on all dimensions (2) Cognitive, (3) Emotional and (4) Social (5) Students' overall experience of speaking with ECAs (three dimensions)	(1) Our calculation of the speaking experience based on all dimensions (2) Cognitive, (3) Emotional, and (4) Social (5) Students' overall experience of speaking with ECAs (three dimensions)	-Views on learning speaking and views on speaking English -Teaching dimension of the experience in the SDS -Overall experience of Enskill (all dimensions included, as rated by the students)	-Overall experience (all dimensions included, based on emoji ratings, 10 time-points)

The analysis of the students' initial experiences of speaking with ECAs, based on the quantitative data (logbook and Q2 ratings), was conducted using descriptive statistics. This was combined with qualitative data from the open-item comments (Q2) and logbook reflections, which in line with were analysed using reflective thematic analysis (Braun & Clarke, 2006; 2019), still yielding the three dimensions cognitive, emotional, and social. The tags were balanced separately per speaking session (the original number of tags divided by the total number of logbook entries).

As suggested by Pallant (2011), four codebooks were created, one for each of the four questionnaires (Q1–Q4). The variables were given names associated with what they measured and guided by the framework's dimensions (applicable in Q2–Q4). Scales were created by means of several indicators (approximately five), e.g., the teaching dimension. The items included were typically not analysed in isolation because they are indicators of larger constructs (DeVellis, 2011). The scale reliability was tested using Cronbach's Alpha, and satisfying alpha values were shown. Typically, the same constructs were measured at two or more time-points.

To explore changes in students' educational experiences, the data was analysed by a wide range of statistical tests in SPSS 28, considering the measurement levels, normality, and sample size (Field, 2018). Considering the small sample size ( $N = 20$ ), data was mainly analysed by means of non-parametric tests, such as the Mann-Whitney and Spearman's correlation. When two sets of scores are drawn from the same entities (here, students), the Wilcoxon signed-rank test was used. Non-parametric tests do not assume a normal distribution, and instead of analysing mean differences, non-parametric tests rank the data so that large scores are represented by high ranks, whereas small scores have low ranks (Field, 2018). Means and standard deviations were used to give the central tendency, and median and range were given to provide measures for nuances. Spearman's correlation test was run for robustness checks to validate the correlation between two different measurement methods (e.g., post-trial rating of variables in Q4 and emoji rating in the logbook). To shed light on group differences, subgroups were used, based on speaking proficiency and gender. Additionally, visualisations of tags and emoji ratings in LoopMe Analytics (LoopMe, 2023) were applied to obtain a data overview and for calculations.

## 7.5 ETHICAL CONCERNS

Regarding the data production methods chosen for this thesis, key ethical principles have been respected (Dörney & Taguchi, 2010; Swedish Research Council, 2017). Guidelines for ethics in Internet research (Franzke et al., 2019; Markham & Buchanan, 2012) were considered because the object of study is connected to human beings and their experiences of a speaking activity in an online SDS, which generated metrics, and their

data from online questionnaires and logbooks was also retrieved. Before starting Studies 2–4, informed consent (Franzke et al., 2019; GDPR, 2018) forms were signed by both the participating students and their legal guardians. In addition, oral consent received by the headmaster and teacher at the selected school. The students were informed that they could withdraw from participation at any time. The links to students' identities were removed before their log-in process, so only the researcher (myself) could link the data from their speaking activities and performances in the SDS to the rest of the collected data. All the results were disconnected from students' and teachers' personal information before publishing to ensure that they cannot be identified. Reflections and “slow thinking” in line with Edling (2019) have been applied for ethical concerns throughout this thesis.

To achieve the aim of this thesis, there was no need for the companies providing the selected SDSs to have any personal information about the students, so another anonymous login solution was arranged to avoid connecting each student's email address to their individual results. All data from the conversations was collected within the software, and the voices were recorded, but without being linked to individual students for privacy protection (GDPR, 2018). Additionally, another concern arises because the company behind the SDS records the students' voices and collects their produced data, including spoken dialogues, performance, and system activity. These extensive datasets may be subjected to data mining (Angeli, 2017), raising data-privacy concerns (GDPR, 2018; Franzke et al., 2019).

Hence, discussions about students' privacy were held with the legal counsel of the University of Gothenburg before the studies began. A consensus was reached about the upholding of ethical standards, and no ethical approval was needed at that time for the studies conducted for this thesis (The Swedish Ethical Review Authority, n.d.; Ethics Review Act, 2003). Additionally, one service agreement was approved by a legal counsel at the University of Gothenburg and signed by the Head of Department, together with the President and CEO of Enskill and LoopMe, including a plan for data storage compliance (GDPR, 2018). The data generated for this thesis has been stored in accordance with the management of research data of the University of Gothenburg (2020). This data will be archived and available after the thesis is finished for data security (Franzke et al., 2019).



In line with the Vancouver recommendations (ICJME, 2023) and to retain my independence as a researcher, no co-writing was performed with the companies behind the SDSs during the process of this thesis. The co-authorship of Papers I–V was transparent, as declared initially in this thesis (see *Distribution of Work*). In contrast to many other studies applying conversational AI in L2 education, as highlighted by Bibauw et al. (2022) and Holmes (2022), this thesis claims to be non-commercially driven. There were no conflicts of interest with the companies behind the selected systems, and permission to use their products and photos was given.

It is worth mentioning the risk of economic interests (Williamson, 2016) in light of the fact that, in 2020, educational technology on the global market capitalised 16.1 billion US dollars, of which 0.8 billion dollars was invested in Europe (EdSurge, 2021). The researcher's responsibility in relation to these economically-driven educational technology companies is to conduct non-biased studies that adhere to ethical principles. To validate the research, it is also important to be transparent about the role of the researcher, which is outlined next, together with the role of the participating teacher (Studies 2–4).

## 7.6 THE ROLE OF THE TEACHER AND THE RESEARCHER

The role of the teacher in Studies 2–4 was to plan the trial periods with the researcher (myself) to fit the lesson plan, available time, and the curriculum and ensure that the studies were carried through accordingly. The teacher introduced the SDSs and the logbooks to the students and talked about the aim of the study, together with the researcher (myself) in Studies 2 and 3, but alone in Study 4 due to the COVID-19 pandemic situation. The teacher was not involved in the execution of the learning activities within the SDS but was on hand to help out with any technical issues, clocking time, and the pedagogical framing of the activity. The teacher enabled approximately 15 minutes of practise in the SDS per lesson, during which the students were supposed to individually practise speaking in English with the ECAs, use the supportive features of the system, and then finish with five minutes of writing reflections about their experiences

of the individual practice of speaking in the SDS. The teacher tried to solve problems and communicate closely with the researcher about any upcoming technical and practical issues during the studies. The teacher was not involved in the analysis or writing process of papers. The selection of SDSs, idea, design, and realisation of the studies, as well as the majority of the analyses, was conducted by the researcher (myself).

In Study 4, the lesson plan for how to use and integrate the SDS in students' practising of speaking English was created in close cooperation with the teacher and in line with the suggested teaching–speaking cycle (Goh & Burns, 2012, p. 153). It was designed in a holistic manner and with a TBLT approach (Ellis et al., 2020, Long, 2015). A series of lessons based on three common everyday-life themes (music, travel, and food) of interest to the students was planned to take place over a period of four months, approximately evenly distributed in time, including breaks for holidays. These themes were chosen in relation to the teacher's interpretation of the curriculum and the available simulations in the SDS Enskill at levels A1 and A2.

## CHAPTER 8

# SUMMARY OF PILOT AND PAPERS

This chapter provides a brief summary of the Pilot and each of the five papers (Papers I–V) included in this thesis. It reports on their key findings, which are grounded in the paper-specific RQs, designed to address the thesis’ three overarching RQs.

### 8.1 PILOT

The goal of the Pilot was to explore if there was any need for or pedagogical interest in conversational AI to be used as a complement to today’s methods provided for students to practise L2 speaking skills in institutional L2 education. The results were not published, but guided the design of Study 1, helped in the pre-testing and refinement of the teacher questionnaire, and were compared with the corresponding findings from Study 1. The participating Spanish teachers ( $N = 9$ ) had a lot of experience and valuable ideas regarding both the implementation and use of digital tools in their language teaching. The results, based on their ratings and open-ended items, echoed previous research that L2 speaking is complex and needs a lot of practice to learn and develop, something described as difficult to provide in a meaningful way in the classroom (Goh & Burns, 2012; Timpe-Laughlin et al., 2020). The results also indicated that digital tools were least frequently used for the teaching of speaking skills. However, by the teachers speaking was considered the most challenging skill to both teach and learn. The results of the Pilot provided a sufficient basis for the decision to carry on with Study 1 as planned, with a larger sample of L2 teachers, and start planning for student studies. Additionally, because conversational AI was introduced through Enskill and ImmerseMe, these two SDSs were tested by the teachers and found to be qualified to employ in the forthcoming studies with students (Studies 2–4) to explore methods of facilitating the teaching of speaking.

## 8.2.PAPER I

Ericsson, E. Sofkova Hashemi, S. & Lundin, J. (2019). Teachers' attitudes and experiences of digital tools within speaking proficiency in foreign teaching and learning – Interactional competence left in the dark. In *Proceedings of ICERI 2019, 12th annual International Conference of Education, Research and Innovation* (pp. 861-870). IATED. [https://www.researchgate.net/publication/338106874\\_TEACHERS'\\_ATTITUDES\\_AND\\_EXPERIENCES\\_OF\\_DIGITAL\\_TOOLS\\_WITHIN\\_SPEAKING\\_PROFICIENCY\\_IN\\_FOREIGN\\_LANGUAGE\\_TEACHING\\_AND\\_LEARNING\\_-\\_INTERACTIONAL\\_COMPETENCE\\_LEFT\\_IN\\_THE\\_DARK](https://www.researchgate.net/publication/338106874_TEACHERS'_ATTITUDES_AND_EXPERIENCES_OF_DIGITAL_TOOLS_WITHIN_SPEAKING_PROFICIENCY_IN_FOREIGN_LANGUAGE_TEACHING_AND_LEARNING_-_INTERACTIONAL_COMPETENCE_LEFT_IN_THE_DARK) [Full paper]. <https://doi.org/10.21125/iceri.2019.0272> [Abstract]

Paper I addresses the situation of the teaching of speaking in Swedish L2 education. It explores teachers' views on and experiences of teaching and using digital tools in their L2 teaching, with a focus on speaking skills (RQ1). Through a multi-method approach, this study serves as an introduction to the pedagogical context of the problem, which is the lack of everyday-life speaking opportunities for students to help them learn and develop speaking skills in a foreign language. The following two RQs were answered in this paper:

1. What are language teachers' attitudes towards digital tools in foreign-language education, especially focusing on speaking proficiency?
2. How do language teachers use digital tools when teaching foreign languages, and especially speaking skills?

Data was gathered from a group of L2 teachers ( $N = 36$ ) with an age range of 32–69 years, who answered a questionnaire. Everyone was qualified and experienced, and most of them taught secondary-school languages such as English, Spanish, French, or German. Quantitative data from the teachers' ratings on a five-point Likert scale and qualitative data from their answers to open-ended questions on the questionnaire related to RQ1 were analysed using descriptive statistics.

This paper reports on gaps that were identified in relation to practices, research, and policies. Firstly, the teachers reported speaking as the most difficult skill to provide opportunities for students to practise, especially in interaction. L3 teachers experienced this as even more difficult than L2 teachers. Some obstacles identified by the whole group of teachers were: students' low self-confidence, large groups, low willingness to speak in class, and lack of access to appropriate material. Although the teachers reported positive views of digital tools for supporting the teaching and learning of the four language skills, speaking was identified as the least digitally supported skill, especially the subskill of interaction, which was "left in the dark". Secondly, the results indicated a gap between the current language education in terms of teaching speaking skills in the classroom and the existing educational technology for language learning that is claimed to engage learners in authentic learning situations and facilitate the learning and development of speaking and listening skills (Bajorek, 2018; Golonka et al., 2014; Johnson, 2019a). The third and final gap was between the relative lack of teachers' attention given to speaking skills and interaction compared to other skills in L2 education and national and international educational policies highlighting the importance of speaking skills and interaction for language learning and development. The teachers' reported experiences of using digital tools were associated with individual factors, such as interest and knowledge about the educational technology, previous experiences, attitudes towards digitalisation, knowledge about previous related research, and the availability of suitable tools and Wi-Fi.

In summary, the paper's participating teachers, who reported interest in and positive attitudes towards digital tools, provided empirical evidence in favour of introducing conversational AI as a complementary method for providing opportunities to practise L2 speaking. The results of the students' experiences are presented in the following paper summaries, which report on Studies II–IV.

### 8.3 PAPER II

Ericsson, E., Sofkova Hashemi, S. & Lundin, J. (2023). Fun and frustrating: Students' perspectives on practising speaking English with virtual humans. *Cogent Education*, 10(1).  
<https://dx.doi.org/10.1080/2331186x.2023.2170088>

Paper II addresses how lower-secondary students ( $N = 25$ ) experienced practising L2 speaking skills by engaging with the AI-based, low-immersive SDS, Enskill. This SDS incorporates animated virtual humans as ECAs for simulated everyday-life conversations around tasks to solve in English, such as ordering food in a restaurant. Through a multi-method approach, the paper provides insights into the students' experiences of practising in the system in relation to learning and developing L2 speaking. The following two RQs were answered in this paper:

1. How do students in lower-secondary school practise speaking English in an SDS with virtual humans?
2. What factors can be identified as the basis for any variations in the group's overall experience?

To explore the students' experiences, data was produced through self-reports in two questionnaires, digital logbooks, semi-structured interviews, and system-generated metrics (showing students' fulfilled learning objectives, simulation, time spent, and mastery scores). The metrics and questionnaire ratings were analysed using descriptive statistics. Data gathered from open-response items in the post-trial questionnaire, written logbook reflections, and transcripts from interviews were analysed in two ways: (i) using reflexive thematic analysis (Braun & Clark, 2006, 2019) of the students' experiences, and (ii) by exploring variations in the overall experience within the group to identify possible patterns and subgroups.

Paper II presented background information about the students, who were accustomed to engaging with digital devices for eight hours on a daily basis and found learning English generally easy and fun. However, 25% of the students were dissatisfied with their opportunities to practise speaking

English in class. Lack of confidence and anxiety were reported as obstacles to their willingness to speak.

The findings from the students' reported experiences of their simulated speaking practice revealed that, on average, in terms of the learning and development of speaking skills, the students found it good, easy, fun, safe, more comfortable, and less stressful using the system than speaking with a human being, but sometimes frustrating due to not being understood. The findings show that the students engaged in conversations with virtual humans in an SDS in order to practise their speaking skills, especially pronunciation and dialogue, in scenarios that were not always relevant to their everyday lives. The thematic analysis of the open-response items yielded seven themes (see Table 8.1.) which, combined with the descriptive statistics, enabled a deeper understanding of the students' experiences. Factors that were suggested to underly the varied levels of frustration experienced by the students include technical issues and constraints with the system, such as not being understood or heard as expected by the virtual humans, or their repetitiveness. Large individual differences were also reported in the aspect of humanising the virtual humans when interacting orally, with some students referring to the virtual humans with personal nouns or human names.

**Table 8.1**

*The seven themes emerging from the reflexive thematic analysis in Paper II*

<b>Speaking Practice Method</b>	<b>Content of the Scenarios</b>	<b>Technology and Design of the SDS</b>
Easy, fun, and innovative	Quite realistic scenarios with a preference for variation and age-adapted tasks	Frustration when not being heard or understood by a limited system
Learning to speak, listen, and read		Human-like vs non-human-like interaction
Less nervous and more comfortable		Suggestions

Additionally, three subgroups of students were found to be clustered based on their self-reported overall experiences (positive, neutral, and negative) in the SDS. These ratings were mirrored in their engagement with the system in terms of time spent on tasks and fulfilled learning objectives related to all the stated learning objectives. The findings indicate higher ratings in the positive group than in the middle and negative groups. For instance, students in the negative group showed less patience with technical issues than the other students, and reports of being bored and frustrated dominated their overall experience of the whole trial. Suggestions for improvements to this kind of system included to relate the content even more to the age group's interests and lives, as well as to provide alternatives with global English variants.

#### 8.4 PAPER III

Ericsson, E. & Sofkova Hashemi, S. (unpublished manuscript).  
(Dis)engagement and usefulness: Students' experiences of practising speaking German in a virtual 360-degree environment.

Paper III addresses how lower-secondary students ( $N = 41$ ) experienced their practise of speaking German in ImmerseMe. The ECAs consisted of pre-filmed human beings in a culturally authentic setting within a virtual 360-degree environment, and the sessions included tasks for the students to solve with them in everyday-life scenarios; for instance, ordering food in a restaurant. This paper provides insights into this way of practising speaking German in an institutional, educational context. Data was produced from students' self-reported experiences in web-based logbooks and questionnaires and complemented by interviews with some students and the teacher. This speaking activity was included as a supplementary speaking activity in their German classes. The following two RQs were answered in this paper:

1. How did the students experience practising pronunciation, the production of utterances, and constrained interaction in everyday-life scenarios with pre-filmed native speakers?
2. How did the students experience being challenged and scaffolded by the system in relation to their learning and development of speaking skills?



Background information concerning the participating students linked the results to pre-trial questionnaire ratings, revealing that they used digital devices for seven hours daily and were positive about using educational digital tools to enhance their speaking skills. The students found learning German fun, but very few were practising speaking outside school.

Descriptive statistics were applied to the students' post-trial ratings to identify how they experienced the speaking sessions regarding the subskills of speaking (pronunciation, the production of utterances, and constrained interaction). Slightly less than half of the students rated the scenarios as highly or very highly realistic. The findings showed that they mostly experienced this L2 speaking practice as good, easy, fun, and safe in relation to the development of speaking skills, especially pronunciation and dialogue, while feeling better prepared and confident for speaking in real life. A few students complained of a bad or boring overall experience with the system.

Additionally, reflective thematic analysis was applied to the students' open-item answers from the post-trial questionnaire and interview transcripts to identify themes based on the students' self-reported experiences in the system in order to identify aspects with bearing upon their experience. As displayed in Table 8.2, the study yielded seven themes indicating a spectrum of engagement and its opposite, disengagement, and another five themes represented within the overarching category of usefulness for the learning and development of German. For instance, the findings showed that the students expressed appreciation of the movement possibilities in the virtual 360-degree environment for the feeling of physically being there and feeling immersed, but argued for a preferably more culturally and interactively engaging environment. The students reported learning and using new words without always understanding their meaning. They were not always able to make connections and construct meaning within the social and cultural context. While practising with the ECAs, the students experienced having a "private teacher" in that the ECA gave them instant feedback. The students found it useful to be independent and concentrated in their self-regulated learning in the system, with possibilities of regulating the supportive features and hence the level of challenge, as well as opportunities for revision and repetition. However, many were annoyed and bored when they were not able to finish the tasks properly and blamed the ECA's inability to hear and understand. In cases of technical issues, there was little patience identified.

**Table 8.2**

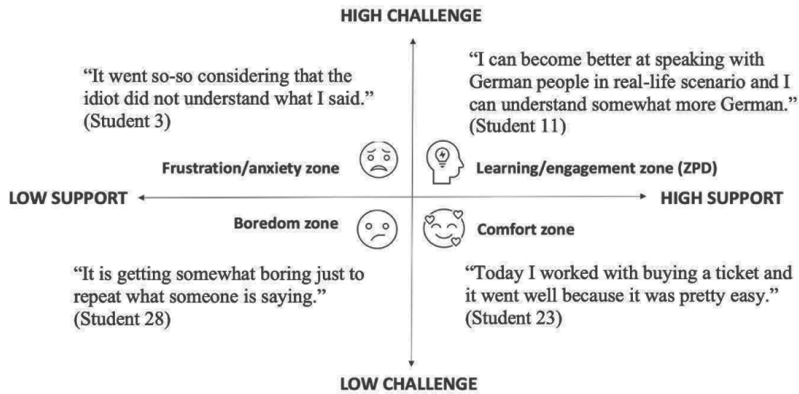
*List of themes emerging from the thematic analysis in Paper III, organised into two overarching categories*

<b>(Dis)engagement (with the system)</b>	<b>Usefulness (for learning and development of L2 skills)</b>
Degree of enjoyment and realism	Improved speaking skills and gained confidence
Varied content	Learning new words without understanding their meaning
Impressions of CAs	Personal teacher with instant feedback
Degree of comfort and safety	Being non-distracted, independent, and self-regulating
Being able to finish dialogues and tasks properly	Revision and repetition
Degree of challenge	
Time and frequency of practice	

Paper III contributes the insight that, although they were immersed in the environment, these students experienced a variety of engagement modes, from continuous engagement in fun, easy, and useful speaking activities to gradual disengagement, experiencing boredom and finding the practice monotonous. This indicates various levels of persistence to continue engaging in the spoken interaction with the ECA, which were linked to the degree of challenge and support experienced, as illustrated in Figure 8.1. Occasionally, when students experienced being too challenged in combination with a level or mode that was too difficult, with too little enabled support from the system, they changed level or mode or sometimes gave up. Expressed boredom was found to be linked to those students who reported experiences of lacking support or challenge.

**Figure 8.1**

*Visualisation of the challenges experienced in ImmerseMe in relation to experiences of support provided*



*Note.* This adaptation is based on a model drawn from Mariani (1997) and Gibbons (2015).

One implication of the findings is that this kind of virtual 360-degree environment, with pre-filmed native speakers as ECAs, would need individual and pedagogical framing when embedding this type of SDS in classroom practice because there are large individual differences between students and some of them might need a guiding hand to find the right balance in their engagement zone while practising in the system. Also, these findings point toward suggested novel approaches for using this kind of system more effectively. These could involve higher interactivity for vocabulary learning and cultural aspects in scenarios with fewer constraints in the interactional design of the system, and more age-adapted content.

## 8.5 PAPER IV

Ericsson, E., Lundin, J., & Sofkova Hashemi, S. (2023). From deadpan machine to relating socially: Middle-school students' experiences speaking English with embodied conversational agents.

Published in *Journal of Research on Technology in Education*.

<http://dx.doi.org/10.1080/15391523.2023.2213902>

Paper IV addresses students' experiences of speaking English face-to-face with ECAs,<sup>6</sup> which are animated virtual humans used for simulated everyday-life conversations around tasks to be solved in English, such as interviewing someone. Through a multi-method approach, the aim of this initial part of the longitudinal Study 4 was to expand the understanding of how middle-school students experience the use of conversational AI for spoken interaction. Swedish 13–14-year-old students ( $N = 22$ ) studying L2 English interacted in simulated everyday-life conversational scenarios in Enskill for 15 minutes during two speaking sessions spread over two weeks. The following RQ was answered in this paper:

1. What characterises middle-school students' experiences when practising speaking with ECAs in a foreign-language<sup>7</sup> learning context?

By implementing the framework developed by the researcher, FoSCAI (see Section 6.4), the students' speaking experiences were dimensionalised into cognitive, emotional, and social aspects, which were also analysed both separately and merged. Background information about the students was collected through a pre-trial questionnaire (Q1), and showed that the majority were positive about speaking English and found it easy and fun, although half of them were nervous about speaking English during lessons. Data about the speaking experiences in the system was produced through rankings of emojis, tags, and open-ended response items that were recorded in the logbook LoopMe immediately after speaking sessions, as

---

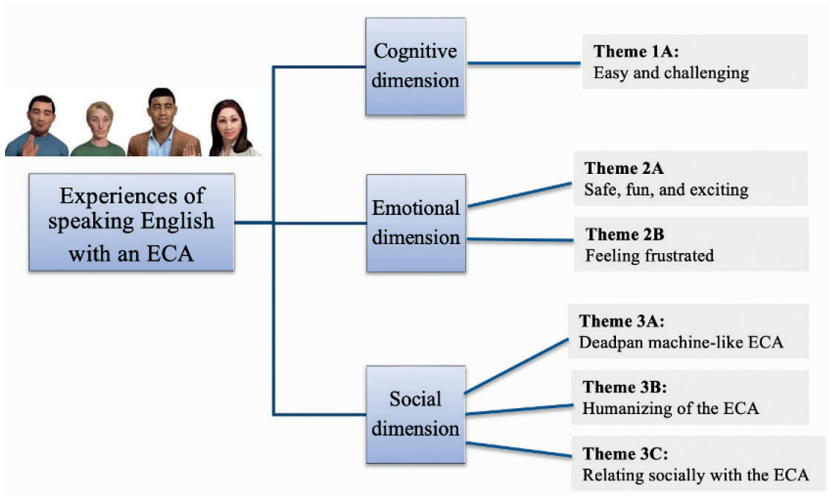
<sup>6</sup> The term ECA was used in this paper, referring to animated characters labelled virtual humans in Paper 2 and in this thesis.

<sup>7</sup> The term foreign language is a synonym for L2, as used in this thesis.

well as ratings and open-ended items in one questionnaire (Q2). Descriptive statistics were applied to the quantitative data from the rankings, combined with reflexive thematic analysis of the qualitative data from the open-ended responses. Data was coded manually within the three dimensions, thereby giving the frame for the themes. To identify and analyse the themes, the qualitative dataset was read through and interpreted using a data-driven (bottom-up) approach, combined with a theoretical interest-driven (top-down) approach, all being further reviewed and refined repeatedly, although still within the three dimensions. The analysis yielded six core themes, as presented in Figure 8.2.

**Figure 8.2**

*The outcome of the thematic analysis. Six themes about speaking English with an ECA*



*Note.* Picture of virtual humans as ECAs, © Alelo Inc. Used with permission.

The findings demonstrated that the students were cognitively satisfied and emotionally and socially engaged when speaking with the ECAs. The students reported experiencing speaking with ECAs as fun and meaningful in terms of the learning and development of speaking skills in English, and they found the ECAs easy to listen to, understand, and speak with while feeling safe. The reported frustration was often associated with the constraints of the virtual human and the system, and most individual differences were found within the social dimension. This social dimension of experiences represented a whole spectrum, ranging from finding the ECAs socially distant, a deadpan machine, to humanising them by ascribing them human attributes, feelings, and behaviours, to even being able to relate to the ECAs socially. The most frequently reported signals that students rated the ECAs as displaying during their interaction were inhuman, deadpan, neutral, and happy. Most students reported feeling comfortable or natural in the conversation with the ECAs and experiencing collaboration with them around the tasks to solve. The majority of these students' overall experiences of practising English in the system were generally average or above, and increased throughout the study, while the levels of frustration decreased.

Paper IV contributed to this thesis by expanding our understanding of how middle-school students relate to ECAs and their importance for the educational experience in an SDS. The study applied FoSCAI as a methodological and analytical lens, separating the speaking experience from the educational experience by applying refined methods for exploring the three dimensions of L2 speaking with ECAs. Relating socially was the theme showing the widest range of individual variation, which may have implications when introducing these ECAs more widely for interaction as a learning activity in L2 education.

## 8.6 PAPER V

Ericsson, E. & Johansson, S. (2023). English speaking practice with conversational AI: Lower-secondary students' educational experiences over time. *Computers and Education: Artificial Intelligence*. Manuscript under review.

Paper V addresses how Swedish lower-secondary students ( $N = 22$ , 13 males and nine females) aged 13–14 years experienced using an SDS with CAs<sup>8</sup> and their potential for facilitating opportunities to speak English in a formal educational situation over time for the learning and development of speaking skills. Study 4, with its longitudinal design and a dataset mainly taking a quantitative approach, provides insights into students' experiences of practising speaking English with animated humanlike virtual humans as ECAs in Enskill during ten speaking sessions spread over a period of four months (see Section 7.3.6). The following specific RQs guided this paper:

1. How can students' educational experiences be characterised after they have practised L2 speaking in an SDS over time?
2. Can differences in students' educational experiences be observed between gender-based sub-groups?
3. Can differences in students' educational experiences be observed between proficiency-based sub-groups?

This study used FoSCAI as a methodological and analytical lens. Data was produced through two sets of questionnaires (pre- and post-trial, and initial and final questionnaire) and systematic reflections recorded in the digital logbook LoopMe. Completed questionnaires were returned by 20 of the students. The ratings were analysed by applying inferential statistics, using variables and associated scales. The outcome, educational experience, was explored and analysed by means of three measures (overall experience of practising in Enskill, speaking with a CA, and the teaching dimension).

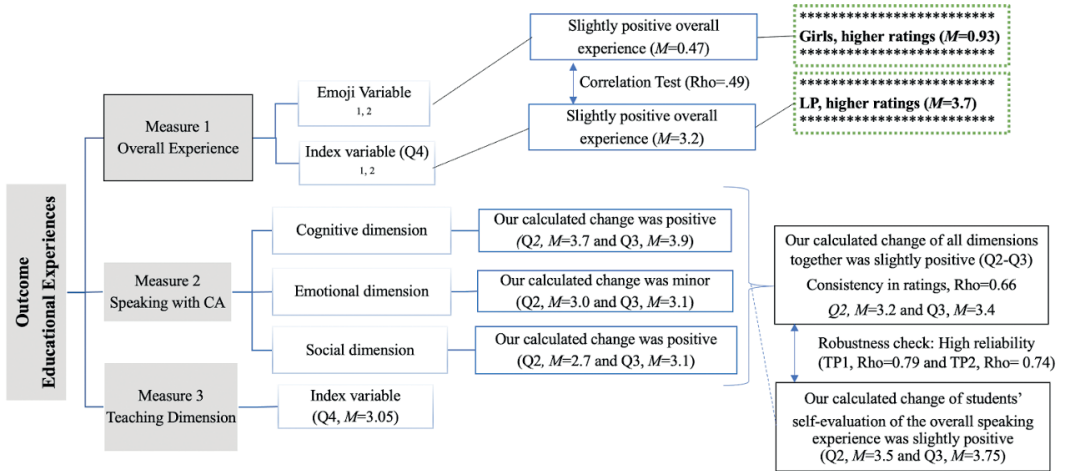
As outlined in Figure 8.3, displaying the statistical results reveals that the various ways of measuring the students' educational experience of practising English speaking skills over time shows a coherent picture of a sustained positive, beneficial trend. It must be noted that the consistency of the students' ratings was quite high between measurement points, as further presented below.

---

8 In this paper, the term CA was used interchangeably with the term ECA, as used in this thesis.

**Figure 8.3**

*Overview of the findings: positive trends for educational experience at the group level, with a statistically significant higher overall experience for female students and the lower-proficiency group of students*



*Note.* Tested for statistically significant differences between subgroups based on

- (i) Low and high proficiency (LP and HP)
- (ii) Gender (female and male)

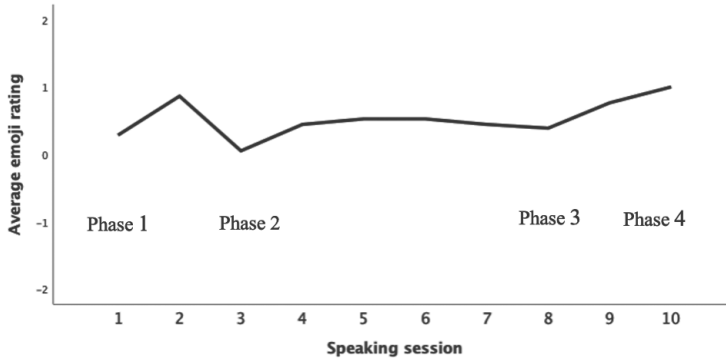
\*\*\*\*\* Green-coloured frame indicates a statistically significant difference.

- (i) **Students' Overall Experience:** At a group level, the longitudinal educational experience indicated mostly positive experiences, sustained during the four-month period. As visualised in Figure 8.4, based on the ratings, four phases were identified that echoed the novelty effect, technical issues, habitual activity, and a final rising trend at the end.



**Figure 8.4**

*Students' overall experience based on the mean of emoji ratings from ten time points, showing a pattern of four identified phases*



*Note.* Likert-scale data from the emojis was coded from -2 to +2, giving 0 as the average value.

When background variables were introduced, the low-proficiency group showed significantly higher ratings than the high-proficiency group on the overall experience (based on Q4 ratings), and female students showed significantly higher ratings than male students on the overall experience (based on emoji ratings).

- (ii) **Speaking with a CA:** The results indicated a positive trend in the speaking experience over time, with higher means at the second measure point (Q3) than the first (Q2), regardless of whether the three measured dimensions were separated or collapsed, as rated by the students in the questionnaires and our calculation of the change in the collapsed dimensions. The results also showed that the students reported speaking with and understanding a CA as easy, and they reported becoming more comfortable speaking with the CA over time.

- (iii) **Teaching Dimension:** The results revealed an average experience for the whole group of students, based on ratings around the mid-point in Q4. These results indicate that the SDS was a helpful way of practising L2 speaking for development in terms of aspects such as receiving feedback from the CA and the system, working with relevant content in the scenarios, and self-regulated learning. The top three rated subskills from those practised were asking/answering questions, interaction/dialogue, and listening.

In sum, Paper V shows that lower-secondary students experienced a sustained educational experience over time, especially low-proficiency students and females. The students experienced being more comfortable in the system and speaking with the ECAs over time. These results suggest that practising L2 speaking in an SDS can be meaningful as a supplementary interaction opportunity in the target language.

## CHAPTER 9

# CONTRIBUTIONS AND DISCUSSION

This thesis aims to explore and gain a nuanced, critical, and comprehensive understanding of how conversational AI is experienced by students during their institutional language education when practising speaking skills through face-to-face interaction. As a departure point for the student studies, language teachers' views and experiences of teaching speaking and using digital tools were explored in Study 1. Then followed Studies 2–4 with lower secondary students, who practised L2 speaking in two separate SDSs and self-reported their experiences over both the short and long term. In line with the iteratively developed FoSCAI, the students' educational experiences were operationalised into cognitive, emotional, social, and teaching dimensions.

The discussion in this chapter first outlines the key findings aligned with the three RQs addressing both teachers (RQ1) and students (RQ2–3), and the methodological contribution of FoSCAI. Then, methodological reflections are presented in relation to the conducted studies, in terms of my role as a researcher, the two SDSs and languages, limitations, self-reporting, trustworthiness, validity, and reliability. The chapter ends by outlining four practical implications for language education and stakeholders and some suggestions for future research. The overarching RQs raised in this thesis were:

RQ1: Focusing on speaking skills, what are language teachers' views and experiences of teaching and using digital tools?

RQ2: How do lower-secondary students experience practising L2 speaking skills in an SDS?

RQ3: How do lower-secondary students experience speaking L2 with an ECA?

## 9.1 LANGUAGE TEACHERS READY TO IMPLEMENT SDS (RQ1)

The findings from the Pilot and Study 1 show that participating Swedish L2 teachers generally reported favourable views of the use of digital tools in their teaching. This is in line with previous studies (e.g., Çelik & Aytin, 2014; Nilsson, 2017), including their views on the potential of implementing SDSs. In contrast with Haelermans (2017), these teachers expressed confidence that they had enough digital competence to support their students when using digital tools in L2 education. They had somewhat less positive views regarding the use of digital tools to facilitate speaking skills than for writing, reading, or listening, which aligns with the fact that they had very little experience of using digital tools for the teaching of speaking. Speaking was ranked as the most challenging skill to teach, thus echoing previous results (Thiriau, 2017), which is explained by the limited opportunities for students to practise speaking in the classroom (Timpe-Laughlin, 2020) and the complexity of speaking (Goh & Burns, 2012). Additionally, these teachers identified hindrances, such as students' low self-confidence and unwillingness to speak, which is also in line with previous studies (e.g., Papin, 2018). They highlighted big groups, the challenge of providing individual feedback, limited lesson time in relation to the curriculum, lack of appropriate teaching materials, and other practical obstacles, as also shown by Çelik and Aytin (2014). The teachers recognised interaction as the most challenging subskill for students to practise; nevertheless, this was also the subskill where digital tools were the least frequently used, despite encouraging policies and beneficial results being found in previous research (e.g., Bibauw et al., 2019). The use of educational digital tools does not reflect the aim of Swedish educational policies (National Agency for Education, 2022a; 2022b), which highlight the use of digital tools in line with the European framework CEFR (2020, 2022). Both of these stress an action-oriented approach to task-based teaching and learning (Ellis et al., 2020).

These findings make contributions to Swedish L2 lower-secondary education, demonstrating language teachers' readiness to implement SDSs incorporating ECAs in order to provide new, complementary methods for teaching and learning speaking skills, while challenging the traditional

roles involved in teaching them, as previously shown by Haelermans (2017). The results confirm that teachers should be involved in integrating SDS as one possible method of providing meaningful speaking practice for students in the target language (Li, 2017; Timpe-Laughlin et al., 2020). Thus, in line with previous international studies (van Doremalen et al., 2016; Timpe-Laughlin et al., 2020), teachers are not a hindrance to introducing conversational AI into a Swedish L2 educational context. The teachers welcome more knowledge about how to use conversational AI for teaching and learning speaking skills. These findings helped to define the starting point for this thesis to continue exploring students' experiences, and also confirmed the choice of using two publicly available SDSs.

## 9.2 FUN, SAFE, AND FRUSTRATING – PRACTISING L2 SPEAKING (RQ2)

The findings of the thesis' Studies 2–4 show that the participating Swedish lower-secondary students generally self-reported themselves as cognitively, emotionally, and socially engaged in the simulated everyday-life scenarios where they practised their L2 speaking, in both the short and long term, although there were differences detected within and between individuals. The students found practising L2 speaking with ECAs in an SDS to be a good way to develop fluency and to learn in a fun, easy, exciting, and safe way in a low-anxiety environment without feeling embarrassed or judged by others. This confirms the findings of previous studies (Bajorek & Hammond, 2019; Johnson, 2019a; Timpe-Laughlin et al., 2022). The speaking practice was experienced as less stressful than speaking with human interlocutors, in line with Bashori et al. (2020). The lower-secondary students thus experienced this way of practising L2 speaking in an SDS as providing opportunities for human-like social interaction as stressed within SLA by among others, Lantolf et al. (2014), Loewen and Sato (2018) and Mahn and Fazalehaq (2020). These findings are also in alignment with previous research in higher education (Andersson et al., 2008; Ayedoun et al., 2015; Morton, 2012).

The various student experiences demonstrated in the studies should be interpreted in light of the different pedagogical approaches and inter-

actional designs that frame the L2 speaking practice in the two SDSs. In ImmerseMe, which applies explicit constraints on meaning, the students chose and repeated utterances from a list in a drill-based practising of speaking, and students memorised structure-based dialogues (Berti, 2020). In contrast, in Enskill, the students were allowed to produce spontaneous speech within the scenario, which enabled social interaction more similar to human-to-human interaction, in line with previous research reported by Johnson (2019a). The studies thus demonstrated that the dialogues permitted for a varied range of spontaneity in the students' utterances depending on the degree of constraints and interactional design of the systems (Bibauw et al., 2019), which had implications for how much social interaction was enabled.

Overall, most of the students found participating in L2 spoken dialogues with the ECAs to be an educationally useful and beneficial way of practising to develop their ability to speak and maintain a flowing dialogue in the target language. What is more, they recognised opportunities for practising interaction, pronunciation, asking and answering questions, and listening in the target language, as shown in previous studies with students (Bajorek, 2018; Johnson, 2019a; Kim et al., 2021; Timple-Laughlin et al., 2022).

Some benefits that these lower-secondary students reported when practising L2 speaking in the SDS were the opportunity to engage in multi-turn, goal-oriented conversations around a task to be solved in everyday-life scenarios. This finding clearly aligns with the communicative and action-oriented approach within SLA and the TBTL (Ellis et al., 2020; Long, 2015). Students experienced authentic input in conversational simulations that focused mostly on meaning, as opposed to form, which were useful for the development of usable skills in contextualised, real-life learning activities, as suggested by Ellis et al. (2020). However, the content and the tasks to solve were not always related to these students' lives and interests, as found previously in student studies (van Doremalen et al., 2016; Soto et al., 2020; Timple-Laughlin et al., 2020). Consequently, students' suggestions for improving the SDSs involved including content that was more relevant to them, which reflects the fact that these particular systems were not designed for the age group or nationality of students targeted in this thesis. The lack of identification with the scenarios negatively affected

their interactional engagement, illustrating the important role of authentic input (Egbert et al., 2009; Li, 2017) and emotional engagement in language learning (Swain, 2013) through being presented with relevant tasks to solve (Ellis et al., 2020). Nevertheless, overall, practising speaking skills in this way was experienced with enjoyment and excitement.

Another finding from the studies concerns feedback (Loewen, 2012), which was experienced as instant, implicit, communicative, and interactional through the responses of the ECA in the turn-taking when the interaction flowed as expected, or through communication breakdowns where there was a need for recasts, as found in similar studies (van Doremalen et al., 2016; Hassani et al., 2016). The students appreciated having freedom in the production of utterances, and suggested that it would be useful to have the ability for more spontaneous speech along the varied continuum of constraints of the SDSs, including ECAs. However, previous research has shown that systems providing constrained interactions and giving corrective feedback seem to lead to greater learning gains (Bibauw et al., 2022). This creates a pedagogical paradox whereby beneficial L2 education presupposes structure, instruction, and corrective feedback. The students expressed appreciation for a high degree of freedom in their speech in order to practise implicit knowledge. Hence, they valued the ability to speak the target language in authentic social interactions with input, output, negotiation of meaning, and feedback just as in real life, which is also the goal of this speaking activity. The constrained learning situations provided in these SDSs provided might be explained by current technological aspects of the SDS, and also by the fact that measuring fluency in spontaneous speech is challenging and still has shortcomings in accuracy (e.g., Bryant et al., 2019). When the potential for variation in student speech production increases, the predictability of the SDS decreases. Additionally, the simultaneously produced verbatim transcripts in both systems were experienced as a sort of corrective feedback on how the student-produced speech was interpreted by the system (Tai et al., 2022).

Overall, these lower-secondary students reported feeling safe in the virtual environment. This corresponds with similar studies, where higher education students did not fear making mistakes in front of others, or feeling judged, while gaining confidence (Andersson et al., 2008; Ayedoun et al., 2015; Huang et al., 2021; Johnson, 2019a). The low-stakes environ-

ment led to reduced anxiety for the students, and they felt comfortable speaking with the ECAs (Li, 2017) while improving their speaking ability (Hsu et al., 2021). Finally, the students were willing to communicate and felt better prepared for L2 speaking in real life (see Ayedoun et al., 2015, 2020; Jeon, 2022; Papin, 2018).

In concordance with Meri-Yilan (2019), when practising L2 speaking in the SDS, the students became more of their own teacher. They reported being able to self-regulate their learning situation in the SDS and feeling independent, including in time and space (Cucchiarini & Strik, 2019; Huang et al., 2021). The students regulated the difficulty level through the modes and by selecting of adequate supportive features within the systems. In sociocultural theory (Vygotsky, 1978) applied to second-language (L2) learning (Gibbons, 2015; Lantolf et al., 2014; Ohta, 1995), L2 learning as a social process is emphasised. A more knowledgeable interlocutor assists the students in bridging the gap between their current speaking ability and their potential developmental level within the ZPD through scaffolding. The students primarily acknowledged the SDS as providing temporary appropriate support and guidance during their speaking practice, enabling them to complete dialogues and progress to new levels through scaffolding. In line with Gibbons' suggested model (2015, p. 17), the students' experiences showed that the balance between the degree of cognitive challenge in the speaking practice in relation to the scaffolding influenced their engagement and perseverance while practising within the SDS. Additionally, some students experienced boredom after a while due to monotonous and repetitive speaking practice, which was sometimes also experienced too easy.

The findings of Study 4 further indicate that SDSs are more beneficial for low-proficiency students, in line with studies by Kim (2016) and Huang et al. (2008) and with the meta-analysis of Bibauw et al. (2022). They show the most positive effects among beginners and low-intermediate students, who still have high levels of anxiety and lack practice in speaking the target language. These findings suggest that the student–ECA interaction provided conditions for low-proficiency students to become engaged in a meaningful way. They were able to practise without feeling too restricted by the interactional constraints of the system and the ECA, while being challenged in their practice. This contrasts with high-



proficiency students, who had already achieved a more developed level of using the L2 and did not always feel stimulated enough. This might be explained by the fact that low-proficiency students still have a limited ability to express themselves spontaneously, and are still learning through using simpler patterns in the language, and therefore need to interact in a dialogue with clear boundaries (Jeon, 2022). The limited capabilities of ECAs (e.g., McTear, 2020) do not enable high-proficiency students to converse about more advanced subjects connected to utterances spoken previously in the dialogue. However, there is previous evidence that high-proficiency students profit more from similar systems due to their ability to listen and understand the ECA's utterances, while the system's lack of adaptivity to beginners' limited ability to speak in terms of pronunciation, fluency, and accuracy can cause problems (Sydorenko et al., 2018).

In line with Harb et al. (2014), it was found that female students, in comparison to male students, reported higher overall educational experiences in the SDSs, which might be explained by females experiencing higher levels of anxiety about L2 speaking in the classroom (e.g., Granfeldt, 2019; Öztürk & Gürbüz, 2013), and males being more often represented in extramural English-speaking activities involving online games (Sundqvist & Sylvén, 2014).

### 9.3 FROM DEADPAN MACHINE TO SOCIAL INTERACTION (RQ3)

The central activity during the L2 practice in the SDSs was speaking with ECAs in the target language, which temporarily takes the place of a classmate or teacher in a dialogue. Students' experiences of speaking with ECAs were iteratively conceptualised, and in Study 4 they were studied in both a dimensionalised and a collapsed form using FoSCAI. The synthesised findings of this thesis' Studies 2–4 show that participating lower-secondary students self-reported consistent and generally promising experiences of speaking with ECAs, both in the short and long terms, although with individual differences in engagement (Ericsson & Johansson, 2023). The students' longitudinal experiences were sustained, indicating a final positive trend in their rating, which finished higher than at the beginning of

the trial. Most of the students were socially and emotionally engaged when speaking with the ECAs, with individually varying levels of frustration.

The students were sensitive to varying degrees to communication breakdowns due, for instance, to technical issues or the constraints of the ECAs. This was also reflected in individual students' reported disengagement and frustration (see Morton et al., 2012; Johnson, 2019a). When the turn-taking between the student and a CA is not working properly, there is a communication breakdown (Gallacher et al., 2018; Wang et al., 2022). Given that SDSs are not technologically fully robust and still have high ASR error rates, it is difficult to control for these types of unexpected breakdowns (Bryant et al., 2019). Even though ASR itself is beyond the scope of this research, these findings show that the students reported more positively when their conversations flowed than in interactions where they experienced communication breakdowns, in line with Fryer et al. (2019). These results are to be expected, and communication breakdowns are recognised challenges in DB-CALL research (Jeon, 2022; Moussalli & Cardoso, 2020). Most students still reported a significant role for motivation, engagement, and enjoyment in achieving their language-learning outcomes, which is parallel to the findings of Johnson (2019a), Morton et al. (2012), and Timpe-Laughlin et al. (2022). Communication breakdowns were very individually experienced and handled, and as suggested by Swain (2013), these findings again highlight the importance of taking emotions into consideration in L2 education.

Another important factor for the students' experiences of the interaction is the social dimension of their speaking experience with an ECA, which was explored and analysed separately in Study 4. The students' experiences reflected a whole spectrum of individual differences in terms of relating socially and feeling a connection with the ECAs, right through to not relating at all, which can also be recognised in previous studies (Bickmore & Rosalind, 2005; Ebadi & Amini, 2022; Lawson et al., 2021). In line with Reeves and Nass' (1996) findings, ideas about humanising or anthropomorphising might add to our understanding of the sense of social interaction. While practising English with animated virtual humans in Enskill, many students ascribed human attributes to them to varying degrees, and some students even related socially with them, establishing a relationship and collaboration with them, as previously suggested by

Craig and Schroeder (2018) and Walker and Ogan (2016). The students were able to express feelings towards the ECAs during the interaction, and identified human signals displayed by the ECAs during the interactions, such as being nice and human-like, with friendly voices. However, this social dimension revealed large discrepancies between individual students, with some also describing the virtual humans as “deadpan machines”. The students’ experiences with the ECAs as social facilitators confirm Craig and Schroeder’s (2018) suggested design principles, with interface elements such as human voices and conversational gestures to emulate human-like behaviour in the ECAs in order to benefit learning. This is referred to as immersive social experiences (Divekar et al., 2021).

Sometimes the students reported feeling even more comfortable speaking with ECAs than with human beings because they did not experience the same pressure, as previously found (Ayedoun & Tokumaru, 2022; Burden & Savin-Baden, 2019). Having time to think of the answer and being less nervous about making mistakes were presented as positive, compared to human-to-human interaction. The students became increasingly comfortable speaking with the ECA as they spent time in the system. The findings show a similar picture of feeling comfortable and safe in the interaction, regardless of the type of ECA. This means that, in terms of feeling more comfortable speaking L2 in an SDS than in dialogues in the classroom, the students’ experiences were not affected by the type of ECA.

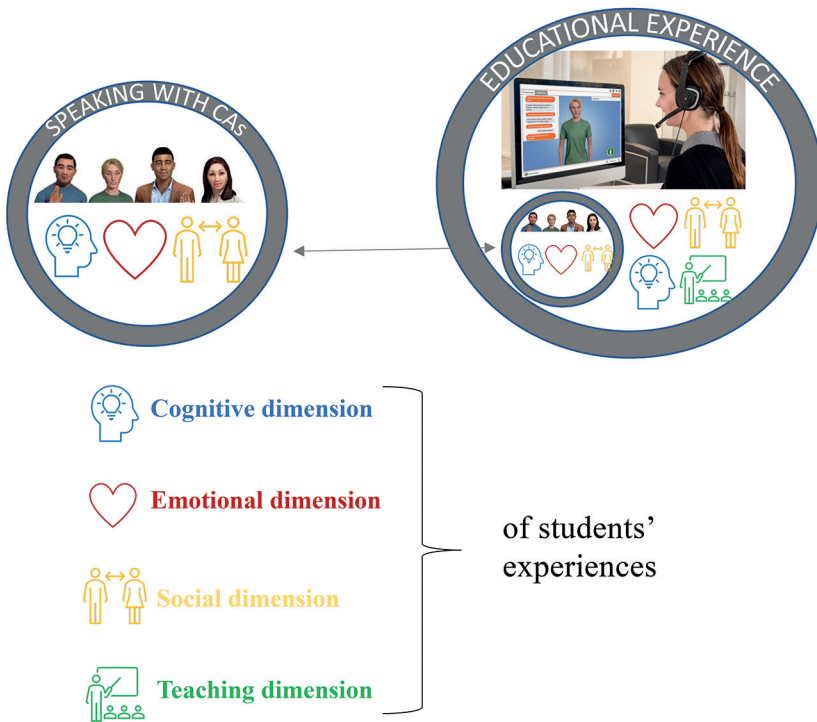
#### 9.4 FRAMEWORK OF STUDENT–CA INTERACTION AND LOGBOOK

This thesis makes methodological contributions in terms of: (i) FoSCAI (see Section 6.4) and (ii) the DB-CALL-adapted instrument in the digital logbook LoopMe (see Section 7.2.4.2). It is also proposed that the latter will be usable for teachers and students in L2 education. This thesis thus contributes to meeting the call for new frameworks to advance theory and practice in the age of AI (Dawson et al., 2023). However, even though the application of FoSCAI and the form of the digital logbook used in this thesis turned out well, and the findings are valid, these methodological contributions are suggestions that were used for the first time in this context and need to be further refined.

- (i) FoSCAI enables research aiming to understand underlying aspects of students' experiences of the student-to-CA interaction (see Figure 9.1). FoSCAI enables refinement of the concept of experience in this online learning situation. Experience is conceptualised as educational experience, including the speaking experience with CAs, which is analysed separately to provide a nuanced understanding.

**Figure 9.1**

*The design of FoSCAI*



*Note.* Dimensionalising students' experiences of L2 speaking with CAs included as the central activity in practising L2 speaking in an SDS in order to gain a nuanced understanding of their educational experiences.

- (ii) The platform-based digital logbook, LoopMe, which was adapted into a DB-CALL study applied in Study 4, and used as a systematic way to follow the students' self-reported experiences in an SDS, is a methodological contribution. This adapted logbook, apart from being applied in similar future research studies, could also be useful to teachers and students when systematically reflecting upon their experiences of the process involved in teaching and learning assisted by, for instance, conversational AI.

## 9.5 METHODOLOGICAL REFLECTIONS

Having a background as an L2 teacher and development manager for digitalisation in education has provided me with a pre-understanding of the Swedish educational L2 context. This can be seen as an advantage in that it provided natural access to both a linguistic understanding of the educational practice and contacts for conducting the studies. It could also be a disadvantage in terms of preconceptions, which, according to Shenton (2004), may be a dilemma when conducting research in qualitative methods. A suggested approach for handling this bias is triangulation (*ibid.*) and collaboration with other researchers during the analyses. Coding was conducted separately for the reflexive thematic analyses (Braun & Clarke, 2019) and collegial review was performed throughout the analysis process including discussions for inter-rater coding reliability between the coders in the few cases where discrepancies arose. In line with Creswell and Creswell (2018), using various data collection methods and analysis techniques within the research deepened the understanding of the object of study.

### 9.5.1 SPOKEN DIALOGUE SYSTEMS AND LANGUAGES

When designing Studies 1–4 of this thesis, two types of meaning-focused SDSs with ECAs were selected, following the set criteria of choosing designs of low-immersive virtual environments that are publicly available. Those two SDSs were Enskill, which provides an animated environment,

and ImmerseMe, which employs 360-degree videos. Neither of the SDSs was designed for this particular age group, or for the Swedish curriculum, which has affected the results. No system enabling spontaneous speech in French, German, or Spanish was found when this research was being initiated, although the initial idea was to practise speaking French and Spanish since the need to practise speaking modern languages (L3) is generally more significant than speaking English (L2) in the Swedish educational context (Aronsson, 2019; Bardel et al., 2019; Erickson et al., 2018; Granfeldt et al. 2021). The selection of these SDSs, including various technical designs, such as the complexity of dialogue management, was considered broad enough to interest the participating students, and this was also confirmed by the teachers in the Pilot. However, this current student sample had English and German as the target L2 and L3 languages, respectively. The scope of which language was being studied was not a priority in this thesis. The focus was on speaking skills, although listening, reading, speaking, and writing were all involved when engaging in speaking practice in both systems.

The empirical data that was derived from the students' speaking activities, such as response accuracy calculated scores, rely heavily on the system's ASR for voice recognition and NLP for interpreting and comprehending diverse speech patterns. However, L2 students may exhibit varying degrees of accents and language variants for which the systems are not adequately trained (Anderson et al., 2008). Previous experiences with SDSs have also indicated bias in ASR, with female users reporting that they feel less well understood by the system (Bajorek, 2019). Healey (2022) emphasises the need for a critical examination of educational digital tools, including privacy and human biases. She questions the notion of a "correct" language variety and the absence of many language variations in existing conversational AI datasets. Another concern about applying conversational AI in an educational context centres on the theoretical and pedagogical ideas integrated into the system (Humble & Mozelius, 2022; Luckin et al., 2018). In this thesis, this is transferred into the feedback and assessment of L2 speaking skills while applying the correct norms of, for instance, fluency and accuracy in the SDSs.

## 9.5.2 LIMITATIONS

In line with the iteratively developed research design, the SDSs were used for approximately one month in each of the first two studies, and then, in an attempt to exclude the novelty effect, Study 4 was designed as a long-term study. Due to the COVID-19 pandemic situation, Study 4 was involuntarily paused for a while in school due to the implementation of distance education. However, the transfer went smoothly, signalling that this way of practising is suitable for both home and classroom settings. Due to the scope of the thesis, the differences between experiencing the practice individually at home or in the classroom in the company of other actively participating classmates were not further investigated. A beta version of the Enskill dashboard was used in Study 2, which caused unexpected technical issues and bugs in the system during the trial, and this affected the number of interrupted conversations. These technical issues with Enskill were resolved before the start of Study 4. However, there were, a few, seemingly random, technical issues occurring throughout all the studies.

It is also important to acknowledge the limitations inherent in the sample size recruited for this study. While the small sample size restricted the generalisability of the findings, it nonetheless enabled an in-depth exploration of the participants' experiences (Levy, 2015). By conducting a qualitative analysis, rich insights and nuanced understandings were gained, which provided a solid foundation for answering the RQs.

It is worth noting that the findings presented in this thesis are valuable in serving as a stepping-stone for further exploration in this emerging area. In order to obtain a more comprehensive understanding of the topic, it is crucial to consider the potential impact of various contextual factors. For instance, conducting the study with a different sample from more diverse schools or regions, or including students with varying socioeconomic backgrounds and levels of digital competence, may yield distinct outcomes (Creswell & Creswell, 2018). These outcomes could potentially be attributed to a range of factors, including the interplay between teachers' classroom management strategies, the digital competence of caregivers, and students' prior experience with technology and digital tools (Krumsvik et al., 2018). Therefore, it is important to recognise that the present findings are contingent upon the specific sample utilised ( $N = 86$ ), and future research should endeavour to explore these aspects across a broader and more diverse scope.

### 9.5.3 SELF-REPORTING, TRUSTWORTHINESS AND INSTRUMENTS

Most of the data production was based on self-reports (Pekruna, 2020), and involved teachers and students providing information about their thoughts, feelings, and behaviours. This method is commonly used in qualitative research and is recognised within CALL research (e.g., Levy, 2015). The use of self-reports may be questioned, however (Dunlosky & Metcalfe, 2008; Craig et al., 2020). This is due to a variety of limitations and biases, as reviewed by Gosling et al. (2004), which might affect the accuracy and quality of data. Humans' retrospective self-judgement of emotions has been considered unreliable in some research, (D'Mello et al., 2008). In Studies 1–4, the teachers' and students' answers were based on subjective interpretations of the questions and their ability to respond to their experiences accurately.

In relation to the students, memory bias was considered during the design process in terms of the combination of methods and the immediate and regular meta-reflection upon emotions in the logbook as a way of eliminating any retrospective unreliability of emotional self-judgement, as suggested by Mills and D'Mello (2014). There was also minimal interference during the studies to avoid response bias or the students feeling obliged to please the researcher or their teacher. In Study 4, the quantitative data produced through students' emoji ratings and tags about their overall experiences helped to identify patterns and tendencies, which the students' qualitative reflections then explained. As suggested by Lackéus (2021), these various ways of producing data positively affect the study's trustworthiness.

However, the extent to which lower-secondary students are able to formulate themselves in open items and the trustworthiness of their ratings is open to discussion. In line with Ejlertsson's ideas about motivating participants (2019), the students in Studies 2–4 proved to be eager to answer questions and self-report. Most of the students frequently participated in the speaking sessions and answered in the logbook and questionnaire, and some students also took part in additional interviews. Based on the analysis, the final digital logbook LoopMe developed and complemented the data generated through the design of the four questionnaires (Q1–Q4). This indicates that the refined logbook might be a potential instrument for future longitudinal studies, avoiding the need for questionnaires, which



are demanding in terms of time and workload for the researcher, and that the students were consistent in their answers and ratings. However, the multi-method approach might have inspired the students to reflect in a different manner, mirrored in nuances in their open-item answers about various aspects of the ECA–student interaction. The questionnaires and digital logbook were found to be feasible and suitable for this age group of students, who used them without any problems, even though the combination of all the instruments was complex. However, if the questionnaires were to be used as a complementary instrument in future studies, it would probably be enough to include fewer items related to the suggested dimensions.

#### 9.5.4 VALIDITY AND RELIABILITY

This thesis combines qualitative and quantitative approaches in order to gain a nuanced, critical, and comprehensive understanding of the underlying factors influencing the students' experiences of an SDS by triangulating different methods and types of data (Creswell & Creswell, 2018). After Studies 1–3, which were mostly qualitatively-based studies employing reflective thematic analysis methods (Braun & Clarke, 2019) combined with descriptive statistics for identifying patterns, Study 4 took an expanded quantitative longitudinal approach (de Vaus, 2001).

In line with principles suggested by Shadish et al. (2002), multi-measurement points in the logbook made it statistically reliable in combination with the questionnaire ratings. Measurement error (de Vaus, 2001) was controlled for using different measures in order to gain higher internal validity (Shadish et al., 2002). The design enabled a robust check for validating the various quantitative measurement methods, which gave close correlation and good consistency, indicating high validity and reliability. Scales were used, with statistical operations for correlation between items for content validity. When computed together, the included variables could provide a good indicator of the dimension of experience. Cronbach's alpha being 0.7 or more was important for the reliability of the computed scales (Field, 2018).

Due to the risk of maturation threats (Shadish et al., 2002), in terms of student fatigue during verbal learning studies, the speaking time was set

to 15 minutes. The self-reported experiences of the practice were explored separately in the various dimensions but also collapsed into one variable, which was validated using the students' own overall rating in the questionnaire. To predict explanations for various self-reported student experiences, independent and latent variables were controlled, which also strengthens the statistical conclusions.

The results that were based on data produced in the questionnaires and analysed using the FoSCAI analytical lens were also validated, using, for instance, the emoji ratings, which were collected via the logbook. Both instruments showed similar results, although the battery of four questionnaires measured experiences in a way that was more complex and time-consuming way for both researchers and students.

## 9.6 PRACTICAL IMPLICATIONS

The following subsections present the practical implications of the key findings that are of importance for L2 teachers and stakeholders.

### 9.6.1 ALIGNMENT WITH SLA PRINCIPLES

As shown in the discussion of findings in relation to RQs 2 and 3, the students' educational experiences were in overall alignment with the established key principles and ideas of SLA (e.g., Ellis et al., 2020; Li, 2017; Nation, 2008; Swain, 2013) which seek to promote effective L2 learning of speaking skills in social interaction (e.g., Lantolf et al., 2014; Loewen & Sato, 2018; Long, 1996). The findings confirm the suggested beneficial role of speech technology in L2 learning and the development of speaking using such technology (e.g., Golonka et al., 2014; Li, 2017; Bajorek, 2018). The participating students recognised opportunities for social interaction in the target language, and stressed the importance of authentic input in a fun, safe, and active way through practising everyday-life tasks, although the content did not always relate to their lives or interests. The findings emphasise the importance of scaffolding in relation to their individual ZPD in order to provide emotional engagement, a low affective filter, and individualised feedback. These findings resonate with insights that engaging in conversations with ECAs in SDSs to practise L2 speaking

has potential for L2 learning and the development of interaction and pronunciation (e.g., Bajorek & Hammond, 2019; Soto et al., 2020; Johnson, 2021). These findings could have practical implications for Swedish L2 education, where conversational AI could facilitate new complementary methods for practising speaking among lower-secondary students. However, these findings, based on students' experiences, contributed knowledge that reveals the complexity of this speaking practice, based on the identified aspects of lower-secondary students' educational experiences of practising L2 speaking with ECAs in SDSs, as presented next.

### 9.6.2 THE COMPLEXITY OF PRACTISING L2 SPEAKING IN AN SDS

Based on the summarised findings of Papers II–V, in practical terms, this thesis reveals the complexity of practising L2 speaking in an SDS. The aspects identified as central to the students' experiences that contribute to their educational experiences signify that the practice of L2 speaking in an SDS is beneficial for learning and development. It must be noted, however, that these aspects are to be interpreted as occurring along an individually differentiated continuum. The aspects can be organised, without any internal ranking, into three main areas, although some aspects interrelate: (i) system functionality, (ii) learning, and (iii) engagement.

**Table 9.1**

*Aspects of students' self-reported experiences in SDSs*

<b>(i) System Functionality</b>	<b>(ii) Learning</b>	<b>(iii) Engagement</b>
Technical issues	Self-regulated	Low-anxiety
Constraints	L2 speaking practice	Comfortable
Properties of ECA	Content of scenario	Enjoyment
L2 interaction		Immersion
		Relating socially to ECA
	Feedback	Support
	Anywhere, anytime	Challenge

For instance, a low level of enjoyment may be experienced as boredom, with nuances in between, and this has a bearing on the student's engagement with the SDS. A low level of relating socially to the ECAs means experiencing them as "deadpan machines", which affects how the learning situation is experienced overall. There are both within-individual and between-individual differences in how this way of practising L2 speaking is experienced. Students' experiences may vary from time to time depending on these aspects, and students also react very differently to various aspects, such as technical issues. Therefore, the pedagogical and individual framing is important, although the SDSs were reported to facilitate self-regulated learning. These aspects may serve as input to enable a pre-understanding for teachers when implementing SDSs in their L2 teaching. These aspects are important to individually varying degrees for how students experience this way of practising L2 speaking, and this can help teachers with their reflective practices in relation to their teaching with conversational AI. Another function is for the developers of SDSs with ECAs to be informed about which aspects students have reported as being important for their educational experience in this kind of system.

### 9.6.3 CHANGES IN THE EDUCATIONAL LANDSCAPE

The findings of this thesis imply that, when conversational AI is introduced, the educational situation will change, and the role of the teacher becomes more that of a facilitator and guide, as suggested previously by Jeon (2022) and Utterberg Modén et al. (2023) in relation to the digitalisation of education. Since students react differently in the SDS, the teacher is still needed to guide and support them when they lack patience, or when they are not being understood as expected by the ECA. It is important to stress that this way of practising speaking would be supplementary to other kinds of educational speaking activities, which means that it is useful to integrate the scenarios and themes into the ordinary lesson plan in order to make them more meaningful. The teacher still needs to be in charge of the pedagogical framing. For instance, practising L2 speaking in an SDS could be seen as preparation for human-to-human interaction in classroom dialogues, using similar content and vocabulary in blended teaching (Grgurovic, 2017).

Furthermore, the implementation of SDSs in language learning provides valuable opportunities for teachers to effectively supervise and evaluate students' spoken interactions, which are typically transient in nature and challenging to assess individually in a classroom setting. The incorporation of SDS dashboards enables educators to utilise learning analytics, as advocated by McNamara et al. (2017), thereby enabling simultaneous monitoring and personalised feedback provision to multiple students. Nonetheless, it is crucial to address concerns surrounding potential surveillance and excessive control over students because their speech and activities are being recorded, analysed, and evaluated within the SDS. These developments prompt a necessary discussion about the balance between monitoring and autonomy. Consequently, this thesis emphasises the ongoing importance of teachers as facilitators and advocates for the establishment of pedagogical framing alongside the integration of educational digital tools and AI in education, thereby ensuring teachers' continued role in providing guidance and support to learners (Hrastinski et al., 2019). The digitalisation of L2 education is represented in this thesis through conversational AI as introduced into the subjects of English and modern languages, which is expected to have the potential to support L2 learning and the development of speaking skills. Because the SDS provides individual opportunities for social interaction in the target language, and presents engaging pedagogical materials, while assessing students' learning progress and giving personalised feedback, the traditional roles of the teacher and student are changed.

#### 9.6.4 SUGGESTIONS FOR IMPROVEMENTS IN SDSs

As a consequence of their experiences with the SDSs, the students and their teacher were eager to suggest ideas for improvement for developers to consider in order to meet the students' various needs and facilitate their L2 learning and development with SDSs. These suggestions have practical implications and could contribute to a more user-centric design and the improved usability of this type of SDS. One crucial aspect that was highlighted was that content should be adapted to the age group and the students' everyday lives, to provide themes they can relate to and find interesting, and scenarios that they may expect to meet later in life. This

was also mirrored in the thesis' findings. The students suggested more variations in the answers and more alternatives to choose from; for instance, when ordering food at a restaurant. They also suggested making greater use of the multidimensional environment, making it more interactive both in terms of specific vocabulary practice and by providing small unexpected dialogues in the scenarios for the practice of interaction. Additionally, there were requests for dialect options for more global English, for instance, to match the Swedish curriculum and involve the target language's culture more actively in the scenarios, both in the settings and in the content of the dialogues. There were also requests for fewer constraints in the turn-taking and more detailed feedback on the students' pronunciation, with corrective personal feedback on individual sounds and stress patterns after dialogues have been completed.

## 9.7 FUTURE RESEARCH

Future research studies are suggested, combined with variables around learning outcomes that are also measured by the system and proficiency tests for a further understanding of the students' self-reports. As a prolongation of Paper V, which lies outside the scope of this thesis, another analysis was conducted on the data produced in Study 4. The findings will be presented in a future paper (in process) about the students' self-assessed L2 learning and development over time, by analysing their speaking in terms of skills, confidence, and views in relation to L2 speaking.

Additionally, students' experiences of the various supportive features provided by the systems could be explored in order to gain insights into which of these may most benefit students' L2 learning and development. This could include features such as transcripts, dashboards, and complementary exercises. Also, from a teacher's perspective, a dashboard in an SDS could support the supervision and assessment of students' speaking activities, which may be useful for the teachers. A dashboard that includes learning analytics of the students' speech makes their spoken performance visible. Due to the ephemeral characteristics of speech and utterances in human-to-human interactions, the assessment of speaking can be challenging and time-consuming (Goh & Burns, 2012). However, the issues of being judged or supervised are then introduced into this "safe environ-

ment”, and this needs further exploration based on experiences from both teacher and student perspectives. It is suggested that the development and improvement of supportive features could be promoted by collaboration between students, teachers, and stakeholders.

The physical setting for this speaking activity is important. In a classroom setting, it would be interesting to include physical and social factors in future studies, as well as in comparison with distance education in students’ private home environments. Further exploration of subgroups, based on aspects such as the proficiency and gender of students, could provide further insights. Reflecting upon the methodology, several lessons can be drawn from the conducted studies about student–CA interaction in SDSs for practising L2 speaking skills in an institutional, educational context. Within the field of DB-CALL, the empirical experiences gained, without an inherent ranking order, can serve as valuable guidance for studies in similar areas:

- Operationalise the concept of experience into the four dimensions of cognitive, emotional, social, and teaching experience, which are all interrelated in the educational experience of L2 speaking practice in SDSs. Analyse the speaking experience with CAs separately and operationalise it into cognitive, emotional, and social dimensions in accordance with the suggested tool, FoSCAI (see Figure 9.1 and Section 6.4). Consequently, the application of FoSCAI is suggested in future similar studies.
- Complement short-term studies with longitudinal studies to shed light on students’ experiences over time in order to limit the possible novelty effects and increase the possibility of measuring learning outcomes. This could eventually also be complemented with retention tests. Control for students’ active time and speaking sessions in the SDSs in order to analyse metrics generated by the SDS as variables measuring performance over time, including in relation to subgroups based on aspects such as grades for speaking proficiency level, gender, or target language. In addition, arrange an experimental design with the use of control groups to establish causality by isolating effects on independent variables (Shadish et al., 2002).
- Study the students’ level of social engagement with the ECAs and its effect on their engagement in the learning activity and measurement of learning outcome in order to gain insights into which components

of the CA seem to be important for the students' educational experience. It might also be useful to include ideas from the revised Agent Persona Instrument (e.g., Schroeder et al., 2021), which focuses on trust in technology, such as the voices of the ECAs.

- Integrate the content of the scenarios into the themes of the lesson plan (this was designed for in Study 4), e.g., food, and potentially also use the scenarios further in real-life human-to-human dialogues in the classroom in blended learning (Grgurovic, 2017). Retention tests could be applied to compare the results with a control group that does not practise the dialogue outside the SDS.
- Study the setting, both virtual and physical, in order to explore the importance of authenticity and immersion in an animated virtual environment and 360-degree videos, and any differences in the effects of the different environments on the educational experience. Explore the real-life physical setting surrounding the learning situation. How does it affect the students' experiences to be practising in the classroom surrounded by other students, alone in a room, or at home?
- Use transcripts of the student-to-CA interaction generated in the SDS in order to learn about students' strategies for communication in interaction with ECAs. These can additionally be used to identify breakdowns in communication and to learn from these how to improve the speech technologies of the SDS. They can also be useful to understand from a pedagogical point of view how students use the target language in dialogues and where the linguistic difficulties are identified, in terms of aspects such as grammar, vocabulary, and pronunciation.
- Further explore the teaching dimension of educational experience in terms of the supportive features of SDSs, such as transcripts of the interaction produced in real-time, in order to gain an understanding of how students use them and the effect of the various support features on the learning outcome and the educational experience.



## CHAPTER 10

# CONCLUSION

The aim of this thesis has been to explore and gain a nuanced, critical, and comprehensive understanding of how conversational AI is experienced by students during their institutional language education when practising speaking skills through face-to-face interaction. This has been achieved through four interrelated studies (Studies 1–4). Initially, in order to answer RQ1, Swedish language teachers were addressed in Study 1 and asked to report on their teaching, especially of speaking skills, and their views on and experiences of using of digital tools. Subsequently, in order to answer RQs 2 and 3, Swedish lower-secondary students were recruited and introduced to conversational AI in Studies 2–4. The participating students self-reported their experiences of practising L2 speaking skills in SDSs, in both the short and long term, during face-to-face interactions with ECAs in simulated everyday-life scenarios. The iteratively developed tool, FoS-CAI, was applied to operationalise these experiences and to gain a nuanced understanding of the students' educational experiences in an SDS, including their experiences of speaking with an ECA.

Concurrent findings from this thesis demonstrate a consensus among the participating Swedish L2 teachers and lower-secondary students regarding the advantages of implementing conversational AI within an educational context to provide supplementary opportunities for practising speaking skills in the target language. Thanks to both the short-term and long-term student studies conducted for this thesis, it can be concluded that the positive educational experiences gained from the speaking practice were not only due to the novelty effect but also remained consistent over time, in terms of both the educational experience within the system and the central activity of speaking with ECAs. The synthesised findings

revealed the students' self-reported educational experiences, such as feeling safe and emotionally engaged, with opportunities for social interaction, mostly in everyday-life scenarios. Consequently, the students' educational experiences aligned to a great extent with Li's (2017) key principles of SLA for effective L2 learning and development. However, sometimes, when there was a low degree of alignment, in terms of relating socially with the ECAs, or breakdowns in communication due to the constraints of the system or technical issues with it, there were reports of students experiencing frustration and boredom linked to their disengagement from the system. Disengagement is also linked to experiences of not being cognitively challenged in relation to support from the system or being unable to relate to the scenarios. These key findings suggest that, when teachers and students use conversational AI in L2 education for practising L2 speaking in social interaction, they still need to be aware of the challenges and limitations that are individually experienced, and that this learning activity still needs pedagogical framing by the L2 teacher.

# CHAPTER 11

## SWEDISH SUMMARY

### 11.1 INTRODUKTION

Den svenska skolans digitalisering har pågått över tid och målsättningar för effektiv användning av digitala lärresurser finns inskriven i både europeiska och svenska styrdokument. Under våren 2023 blev dock den reviderade digitaliseringsstrategin för det svenska skolväsendet pausad av regeringen och det förs en debatt kring digitaliseringens syfte och mål samt i vilken utsträckning den ska fortsätta på samma vis som förut (Beijmo, 2023; Selwyn, 2023b; Thorell et al., 2023).

När det gäller frågan om digitalisering av skolans språkundervisning - som är i fokus i föreliggande avhandling - finns det få studier om tillämpning av digitala lärresurser och mer kunskap behövs. Med utgångspunkt i detta var syftet i avhandlingen att utforska och få en nyanserad, kritisk och omfattande förståelse för hur konversations-AI (McTear, 2020) i en institutionell språkundervisningskontext upplevs av högstadieelever som i interaktion och dialog med en konversations-AI övar på sina muntliga språkfärdigheter. Utöver att bidra med ny kunskap om en lärandesituation där elever använder muntliga dialogsystem för att utveckla sitt mål språk (Bibauw et al., 2019), bidrar denna avhandling även med en konceptualisering av upplevelsen. Genom att studera de utforskade elevernas upplevelser (*experiences*) kognitivt, emotionellt, socialt samt i relation till undervisning via dialogsystemet med fokus på lärande och utveckling av elevens muntliga språkfärdighet i mål språket nås en djup förståelse för utbildningssupplevelsen i systemet.

Utvecklingen av konversations-AI har möjliggjort användning av så kallade förkroppsligade konversationsagenter (på engelska, *embodied conversational agents*, ECA; Craig & Schroeder, 2018), och chatbots (Huang et al., 2021) i olika samhälleliga praktiker där interaktion är ett centralt inslag. Ett aktuellt användningsområde är skolans språkundervisning i engelska och moderna språk med målet att erbjuda eleverna ytterligare tillfällen att öva på sina muntliga språkfärdigheter i en vardaglig kontext. Mer precis att nya möjligheter och tillfällen för eleverna att öva på målspråket i interaktion med konversationsagenter i ett vardagligt sammanhang och där gemensamma uppgifter ska lösas, som exempelvis att beställa mat och dryck på restaurang (Berti, 2020; Johnson, 2019a).

Att tala på ett målspråk kan dock upplevas utmanande, vilket enligt tidigare forskning hänger samman med att tala är en kombinatorisk färdighet som ställer höga krav på till exempel en elev i skolan att kunna formulera sig och att använda explicit kunskap i form av ordförråd, grammatik, uttal och flyt i olika kontexter (Goh och Burns, 2012). I muntlig interaktion ska dessutom samtalspartnerns yttranden uppfattas, tolkas och besvaras i realtid, vilket ställer stora kognitiva, emotionella, och sociala krav på eleven. Forskning har här visat att många elever därav upplever det både svårt och ångestladdat att tala inför andra i klassen, inte minst om det är kopplat till någon typ av bedömning av läraren (Ayedoun, et al., 2015; Timpe-Laughlin, 2020). En annan utmaning här är att eleverna ofta har en begränsad kontakt med målspråket både i och utanför skolan vilket gör undervisningen extra viktig vad gäller möjlighet att öva på språket (Bibauw et al., 2019; Kukulska-Hulme & Lee, 2020). En ytterligare utmaning är stora klasser i skolan vilket medför begränsad tid att som lärare ge feedback och bedöma elevers muntliga språkfärdigheter (Li, 2017). Med denna korta bakgrund besvarade avhandlingen följande forskningsfrågor:

- (i) Med fokus på muntlig språkfärdighet, vilka är språklärares uppfattningar om och erfarenheter av att undervisa och använda digitala läresurser?
- (ii) Hur upplever högstadiel elever att i ett muntligt dialogsystem öva muntlig språkfärdighet i ett målspråk?
- (iii) Hur upplever högstadiel elever att prata ett målspråk med en konversationsagent?

## 11.2 TEORI OCH METOD

Avhandlingens teoretiska inramning utgörs av en sociokulturell syn på lärande och särskilt att en elevs lärande och utveckling av ett målspråk sker genom social interaktion och i samspel med andra mer kunniga individer som kan stötta eleven (Vygotsky, 1978). Med andra ord, facilitera att en elev lär och utvecklar sina muntliga språkliga färdigheter genom en möjlighet att aktivt använda språket i muntlig interaktion och då i vardagsnära situationer som intresserar och engagerar eleven (Li, 2017). Den kommunikativa och aktiva språksynen som präglar avhandlingen tar även sin utgångspunkt i Long's (1996) interaktionshypotes. Denna hypotes föreskriver att en elev utvecklar sin språkförmåga genom (i) språklig input genom det som sägs, (ii) output, (iii) yttranden som han eller hon producerar, (iv) med tydlig direkt återkoppling och (v) att betydelser och innebörder klargörs i det som sägs i interaktionen. Vidare har avhandlingen använt sig av de som Li (2017) benämner i termer av nyckelprinciper för effektivt språklärande, såsom (i) autentisk input, (ii) medvetenhet kring form, (iii) möjlighet till interaktion, (iv) individuell feedback, (v) lågaffektiva filer, och (vi) att det ska finnas en miljö att använda språket i.

I avhandlingens studier med högstadiel elever användes *experience* (på svenska, upplevelse) som analysenhet (Levy, 2015; Roth and Jornet, 2014). Vidare, och genom ett iterativt tillvägagångssätt, utvecklades ett ramverk *Framework of Student-CA Interaction*, (förkortas FoSCAI) för att studera och analysera elevers utbildningsrelaterade upplevelser (på engelska, *educational experiences*) som skapas genom interaktionen mellan elev och konversationsagent. Genom att operationalisera analysenheten *experience* i en kognitiv, känslomässig, social, respektive undervisningsrelaterad dimension bidrog ramverket till att rapportera nyanserade insikter kring en elevs utbildningsrelaterade upplevelse i det muntliga dialogsystemet, i vilken även den centrala aktiviteten att tala med en konversationsagent ingår. Den specifika upplevelsen (på engelska, *speaking experience*) analyserades i sin tur både som uppdelad i tre separata dimensioner (en kognitiv, en känslomässig, respektive en social dimension) samt som en helhet.

### 11.3 DATAANALYS

Avhandlingen använde flera olika metoder för att utforska elevernas självrapporterade *experiences* både över kort och lång tid. Data producerades med hjälp (i) av enkäter, (ii) digital loggbok, (iii) intervju, samt (iv) loggad data kring elevernas aktivitet och resultat från dialogsystemet. Kvantitativ data i form av skattningar och loggad data analyserades med hjälp av deskriptiva och statistiska uträkningar i SPSS (Field, 2018) och med målet att identifiera mönster i data. Kvantitativ data kompletterades med kvalitativ data i form av (i) öppna svar från öppna enkätfrågor, och (ii) elevreflektioner i loggboksform, som analyserades med reflexiv tematisk analys (Braun & Clarke, 2019).

### 11.4 GENOMFÖRDA STUDIER

I avhandlingen rapporteras resultat baserade på fyra empiriska studier, varav en tillhörande pilotstudie med språklärare (lärare, totalt  $N = 45$ ) som genomfördes i samband med språkläraryrkeförningars årsmöten. Här efter följde tre studier med högstadielärover ( $N = 88$ ) som genomfördes som en del av en undervisningssituation placerad i en verklig klassrumskontext i Sverige. I studierna användes två så kallade publikt tillgängliga muntliga dialogsystem med förkroppsligade konversationsagenter, Enskill (Alelo, 2022) respektive ImmerseMe (n.d.). Resultaten från studierna rapporterades i fem vetenskapliga tidskriftsartiklar (Paper I-V).

Den första studien genomfördes med språklärare i Sverige. Målet var att, innan vidare studier med högstadielärover, få en fördjupad förståelse av den pedagogiska kontext som kommande studier skulle genomföras i. Resultaten visade att språklärares syn och erfarenheter av digitala lärresurser i relation till muntlig språkfärdighet pekade på tre glapp i relation till undervisningspraktiken, policy- och styrdokument, samt tidigare forskningsstudier som pekar på nyttan med digitala lärresurser för muntlig språkfärdighet. Vidare framkom att trots att 'tala' uppskattades vara den svåraste färdigheten att både lära ut och lära sig, så var det just i relation till den färdigheten som digitala lärresurser användes minst. Studien visade även att lärarna var positivt inställda till att använda konversations-AI och då även de två utvalda dialogsystemen.

I avhandlingens andra studie utforskades högstadieelevers självrapporterade upplevelser från när de övat att prata engelska med en animerad förkroppsligad konversationsagent i virtuella miljöer i Enskill. Resultaten visade att eleverna upplevde sitt övande som både lätt och roligt, men också som frustrerande beroende på individuella reaktioner i relation till systemets begränsningar och tekniska problem. Resultaten var i linje med de ovan omskrivna nyckelprinciperna för språklärande (Li, 2017) såsom möjligheter för social interaktion utan att känna sig ängslig att prata.

I den tredje studien utforskades högstadieelevers erfarenheter och upplevelser när de i en 360-graders virtuell miljö i ImmerseMe övade på att prata tyska med filmade infödda talare som konversationsagenter. Resultaten visade att eleverna tyckte att aktiviteten var användbar för att lära sig målspråket med bättre flyt och uttal, samt på ett roligt, tryggt och engagerande sätt stärka självförtroendet. Eleverna rapporterade dock att de i olika grad kunde relatera till innehållet i dialogerna och att de inte alltid förstod orden de använde. Därutöver visade resultaten att eleverna upplevde sig vara medryckta i aktiviteterna och i miljön, men att övningsarna ändå kunde upplevas som tråkiga och monotona. ImmerseMe var konstruerat så att eleverna valde exakta formuleringar som de upprepade enligt så kallade skript. Olika nivåer av uthållighet kunde identifieras i analysen av data och därefter kopplas till upplevd nivå av utmaning och stöttning i systemet.

I den fjärde studien utforskades högstadieelevers erfarenheter och upplevelser när de i systemet Enskill, och över en längre tid, övade att prata engelska med animerade förkroppsligade konversationsagenter. Data producerades och analyserades med hjälp av det ovan beskrivna ramverket FoSCAI. Resultaten visade att eleverna mestadels var kognitivt utmanade samt emotionellt och socialt engagerade i interaktionen med konversationsagenten, både över kort och lång tid. Rapporterad frustration hos eleverna kunde kopplas till graden av (icke)humanisering och förmåga att känna social samhörighet med agenten. Eller som några elever uttryckte det - som en "livlös maskin" (på engelska *deadpan machine*).

## 11.5 RESULTAT OCH BIDRAG

Avhandlingens sammantagna resultat visade att lärare och högstadiel elever överlag var positivt inställda och redo att börja använda konversations-AI i språkundervisningen. Resultaten visar vidare att lärare behöver vara medvetna om de muntliga dialogsystemens utmaningar och begränsningar som elever på individuell basis kan uppleva olika. Vad gäller den sociala dimensionen i relation till konversationsagenten framkom en stor variation - eleverna rapporterade alltifrån ingen relation till agenten, till att humanisera och tillskriva agenten mänskliga egenskaper, inklusive en upplevd social relation och ett samarbete. Graden av upplevda känslor och social relation till agenten hade betydelse för helheten hur eleverna upplevde att öva prata målspråket i dialogsystemet. Överlag, föreföll eleverna uppleva att det var en lätt, rolig och bra aktivitet för att hålla igång en dialog på målspråket i en social interaktion situerad i en säker miljö. Eleverna fick öva på att lösa uppgifter i vardagliga situationer, men resultaten visade att de inte alltid upplevdes som relaterade till deras vardag, vilket påverkade engagemanget negativt. Graden av utmaning i förhållande till stöd av dialogsystemets stöttande funktioner hade också tydlig koppling till nivå av engagemang i övningen.

Avhandlingens metodologiska bidrag är ramverket FoSCAI och den digitala plattformsbaserade loggboken vid namn LoopMe som anpassats efter den aktuella språklärandesituationen i ett muntligt dialogsystem (Ericsson, 2022). LoopMe erbjuder systematisk datainsamling över tid och kan förutom för forskare också vara till nytta för lärare för att följa elevernas lärandeprocesser i ett muntligt dialogsystem.

Avhandlingens praktiska implikationer är (i) att elevernas sammanlagda utbildningserfarenheter i dialogsystemen är i linje med SLA-principerna för effektivt språklärande (Li, 2017), vilket pekar på att konversations-AI kan erbjuda gynnsamma möjligheter för elever att öva på att tala målspråket i social interaktion och med autentisk input och direkt feedback, (ii) baserat på avhandlingens sammanlagda fynd identifieras interrelaterade aspekter som karakteriserar eleverna supplevelse i ett muntligt dialogsystem. Dessa aspekter är indelade i tre områden; systemfunktionalitet (exempelvis begränsningar, konversationsagentens egenskaper), lärande (exempelvis innehåll, och feedback) och engagemang (exempelvis låg talängslan,



glädje, och social relation med konversationsagent). Aspekterna visar på en komplexitet som kan tolkas som i ett spektrum vilka i olika grad för olika elever bidrar till gynnsam talövning i dialogsystemet. Att ha kännedom om dessa aspekter och lärandeaktivitetens komplexitet kan underlätta för lärare när konversations-AI ska introduceras i språkundervisningen och för att möta elevers olika reaktioner. Kännedom om dessa aspekter kan även vara till nytta för både utvecklare av muntliga dialogsystem med förkroppsligade konversationsagenter och forskare inom fältet, (iii) införandet av konversations-AI utmanar den traditionella utbildningssituationen i relation till undervisning och lärande av muntlig språkfärdighet. I muntliga dialogsystem reglerar eleverna själva sitt språklärande och sin utveckling, vilket också kan bli synliggjort med formativ och summativ feedback i systemets elev-vy (på engelska, dashboard). Läraren får härigenom mer rollen av någon som underlättar, än aktivt undervisar, men som samtidigt skapar en pedagogisk inramning. Vidare kan lyftas fram att muntliga dialogsystem fångar det annars flyktiga talet som är svårt att ge feedback på samtidigt till en klass, och lärar-vyn kan ge inblick i elevens både muntliga aktivitet i systemet och hans eller hennes resultat. Utifrån avhandlingens resultat ges förslag på vidare forskning där ramverket appliceras och eventuellt förfinas för att fördjupa kunskaperna exempelvis elevers förmåga att uppleva social samhörighet med förkroppsligade konversationsagenter och bakomliggande faktorer som påverkar detta, och hur den sociala dimensionen i sin tur påverkar språkärandet som sker i de muntliga dialogsystemen.



## REFERENCES

- Ai-Lim Lee, E., Wong, K. W., & Fung, C. C. (2010). How does desktop virtual reality enhance learning outcomes? A structural equation modeling approach. *Computers and Education*, 55(4), 1424–1442. <https://doi.org/10.1016/j.compedu.2010.06.006>
- Albright, G., Bryan, C., Adam, C., McMillan, J., & Shockley, K. (2018). Using virtual patient simulations to prepare primary health care professionals to conduct substance use and mental health screening and brief intervention. *Journal of the American Psychiatric Nurses Association*, 24(3), 247–259.
- Alelo. (2022). *Enskill English® Revolutionize English language teaching with AI*. Alelo Inc. Retrieved from <https://www.alelo.com/english-language-teaching/>
- Alès, Z., Dubuisson Duplessis, G., Serban, O., & Pauchet, A. (2012). *A Methodology to Design Human-Like Embodied Conversational Agents*. International Workshop on Human-Agent Interaction Design and Models (HAIDM 12). Valencia, Spain. [https://hal.science/hal-00927488/file/2012-dubuisson\\_duplessis-A-Methodology-to-Design-Human-Like-Embodied-Conversational-Agents-based-on-Dialogue-Analysis.pdf](https://hal.science/hal-00927488/file/2012-dubuisson_duplessis-A-Methodology-to-Design-Human-Like-Embodied-Conversational-Agents-based-on-Dialogue-Analysis.pdf)
- Alm, A., & Nkomo, L. M. (2020). Chatbot experiences of informal language learners: A sentiment analysis. *International Journal of Computer-Assisted Language Learning and Teaching*, 10(4), 51–65. <https://doi.org/10.4018/IJCALLT.2020100104>
- Anderson, J. N., Davidson, N., Morton, H., & Jack, M. A. (2008). Language learning with interactive virtual agent scenarios and speech recognition: Lessons learned. *Computer Animation and Virtual Worlds*, 19(5), 605–619. <https://doi.org/10.1002/cav.265>

- Anderson, T., Rourke, L., Garrison, D. R., & Archer, W. (2001). Assessing teaching presence in a computer conferencing environment. *Journal of Asynchronous Learning Networks*, 5(2), 1–17. <https://doi.org/10.24059/olj.v5i2.1875>
- Angeli, C., Howard, S. K., Ma, J., Yang, J. & Kirschner, P. A. (2017). Data mining in educational technology classroom research: Can it make a contribution? *Computers and Education*, 113, 226–242. <http://dx.doi.org/10.1016/j.compedu.2017.05.021>
- Aronsson, B. (2020). A study of learner profiles in Spanish as a second language in a Swedish instructional setting. *Journal of Linguistics and Language Teaching*, 11(1), 69–90.
- Aronsson, B. (2023). Language learning activities in the Spanish L2 classroom related to a task-based framework: What types are the most commonly occurring according to Swedish learners? *Language Teaching Research*, 0(0), 1–27. <https://doi.org/10.1177/13621688221144282>
- Atkinson, R. K., Mayer, R. E., & Merrill, M. M. (2005). Fostering social agency in multimedia learning: Examining the impact of an animated agent's voice. *Contemporary Educational Psychology*, 30(1), 117–139.
- Ayedoun, E., Hayashi, Y., & Seta, K. (2015). A conversational agent to encourage willingness to communicate in the context of English as a foreign language. *Procedia Computer Science*, 60(1), 1433–1442. <https://doi.org/10.1016/j.procs.2015.08.21>
- Ayedoun, E., Hayashi, Y., & Seta, K. (2020). Toward personalized scaffolding and fading of motivational support in L2 learner-dialogue agent interactions: An exploratory study. *IEEE Transactions on Learning Technologies*, 13(3), 604–616. <https://doi.org/10.1109/TLT.2020.2989776>

- Ayedoun, E., & Tokumaru, M. (2022). Towards emotionally expressive virtual human agents to foster L2 production: Insights from a preliminary Woz experiment. *Multimodal Technologies and Interaction*, 6(77), 1–14. <https://doi.org/10.3390/mti6090077>
- Babbel. (2023). *Language learning that works*. Babbel. [https://uk.babbel.com/?locale=en\\_GB&gl=1\\*rkkbkf\\*gcl\\_au\\*Njg0MTI0ODEzLjE2ODI4NTcxMjU](https://uk.babbel.com/?locale=en_GB&gl=1*rkkbkf*gcl_au*Njg0MTI0ODEzLjE2ODI4NTcxMjU)
- Bajorek, J. P. (2017). L2 pronunciation in CALL: The unrealized potential of Rosetta Stone, Duolingo, Babbel, and Mango Languages. *Issues and Trends in Educational Technology*, 5(1), 24–51.
- Bajorek J. P. (2018). Speech technology for language learning: Research and today's tools. *Laureate-Cambridge Online Language Learning Research Network (OLLReN)*. Cambridge University Press. [https://ollren.org/ld.php?content\\_id=44618256](https://ollren.org/ld.php?content_id=44618256)
- Bajorek, J. P. (2019). May, 10). *Voice recognition still has significant race and gender biases*. Harvard Business Review. <https://hbr.org/2019/05/voice-recognition-still-has-significant-race-and-gender-biases>
- Bajorek, J. P., & Hammond, M. (2019). *Speaking of Language Technology*. [Doctoral Dissertation, The University of Arizona]. <https://www.proquest.com/dissertations-theses/speaking-language-technology/docview/2330609145/se-2?accountid=11359>
- Bardel, C. (2019). *Kommunikativ språkundervisning och en handling-sorienterad språksyn* [Communicative language teaching and an action-oriented language approach, my translation from Swedish]. National Agency for Education. [https://larportalen.skolverket.se/LarportalenAPI/api-v2/document/path/larportalen/material/inriktningar/1a-moderna-sprak/Grundskola/910a-muntlig-sprakutveckling/del\\_02/Material/Flik/Del\\_02\\_MomentA/Artiklar/M1\\_GRGYVX\\_02A\\_01\\_kommunikativ.docx](https://larportalen.skolverket.se/LarportalenAPI/api-v2/document/path/larportalen/material/inriktningar/1a-moderna-sprak/Grundskola/910a-muntlig-sprakutveckling/del_02/Material/Flik/Del_02_MomentA/Artiklar/M1_GRGYVX_02A_01_kommunikativ.docx)

- Bardel, C., Erickson, G., & Österberg, R. (2019). Learning, teaching and assessment of second foreign languages in Swedish lower secondary school: Dilemmas and prospects. *Apples – Journal of Applied Language Studies*, 13(1), 7–26. <http://dx.doi.org/10.17011/apples/urn.201903011687>
- Bashori, M., van Hout, R., Strik, H., & Cucchiarini, C. (2020). Web-based language learning and speaking anxiety. *Computer Assisted Language Learning*, 35(5–6), 1–32. <https://doi.org/10.1080/09588221.2020.1770293>
- Bashori, M., van Hout, R., Strik, H., & Cucchiarini, C. (2021). Effects of ASR-based websites on EFL learners' vocabulary, speaking anxiety, and language enjoyment. *System*, 99, 1–16. <https://doi.org/10.1016/j.system.2021.102496>
- Bates, A. W. (2019). *Teaching in a Digital Age. Guidelines for designing teaching and learning* (2nd ed.). BCcampus. <https://open.umn.edu/opentextbooks/textbooks/teaching-in-a-digital-age-guidelines-for-designingteaching-and-learning-for-a-digital-age>
- Bax, S. (2003). CALL, past, present and future. *System*, 31, 13–28.
- Bax, S. (2011). Normalisation revisited: The effective use of technology in language education. *International Journal of Computer-Assisted Language Learning and Teaching*. 1(2), 1–15.
- Beijmo, M. (2023, May 2). Kriget mot skärmarna är helt meningslöst. Att avdigitalisera skolan är fel väg. [The war against the screens is completely pointless; To de-digitalise school is wrong, my translation from Swedish]. *Aftonbladet*. <https://www.aftonbladet.se/kultur/a/KneKeo/mattias-beijmo-om-skolan-och-digitaliseringen>
- Belda-Medina, J., & Calvo-Ferrer, J. R. (2022). Using chatbots as AI conversational partners in language learning. *Applied Sciences* (Switzerland), 12(17), 8427. <https://doi.org/10.3390/app12178427>

- Bellegarda, J. R. (2014). Spoken language understanding for natural interaction: The Siri experience. In: J. Mariani, S. Rosset, M. Garnier-Rizet, & L. Devillers, (Eds.), *Natural interaction with robots, knowbots and smartphones*. Springer. [https://doi.org/10.1007/978-1-4614-8280-2\\_1](https://doi.org/10.1007/978-1-4614-8280-2_1)
- Bergdahl, N., Nouri, J., Fors, U., & Knutsson, O. (2020). Engagement, disengagement and performance when learning with technologies in upper secondary school. *Computers and Education*, 149, 103783. <https://doi.org/10.1016/j.compedu.2019.103783>
- Bergman, M. M. (2008). The straw men of the qualitative-quantitative divide and their influence on mixed methods research. In M. M. Bergman (Ed.), *Advances in mixed methods research theories and applications* (pp. 11–23). Sage.
- Bergmann, K., & Macedonia, M. (2013). A virtual agent as vocabulary trainer: Iconic gestures help to improve learners' memory performance. In R. Aylett, B. Krenn, C. Pelachaud, & H. Shimodaira (Eds.), *Intelligent virtual agents IVA 2013 Lecture Notes in Computer Science 8108* (pp. 139–148). Springer. [https://doi.org/10.1007/978-3-642-40415-3\\_12](https://doi.org/10.1007/978-3-642-40415-3_12)
- Berti, M. (2019). Italian open education: Virtual reality immersions for the language classroom. In A. Comas-Quinn, A. Beaven, & B. Sawhill (Eds.), *New case studies of openness in and beyond the language classroom* (pp. 37–47). Research-publishing.net. <https://doi.org/10.14705/rpnet.2019.37.965>
- Berti, M. (2020). Learning technology review: ImmerseMe. *Calico Journal*, 37(3), 321–330. <http://doi.org/10.1558/cj.39714>
- Bibauw, S., François, T., & Desmet, P. (2019). Discussing with a computer to practice a foreign language: Research synthesis and conceptual framework of dialogue-based CALL. *Computer Assisted Language Learning*. 32(8), 827–877. <https://doi.org/10.1080/09588221.2018.1535508>

- Bibauw, S., François, T., Van den Noortgate, W., & Desmet, P. (2022). Dialogue systems for language learning: A meta-analysis. *Language Learning & Technology*, 26(1), 1–24, <https://hdl.handle.net/10125/73488>
- Bickmore, T. W., & Cassell, J. (2005). Social dialogue with embodied conversational agents. In: J. C. J. van Kuppevelt, L. Dybkjær, & N. O., Bernsen (Eds.), *Advances in natural multimodal dialogue systems. Text, speech and language technology* (pp. 1–32). Springer. [https://doi.org/10.1007/1-4020-3933-6\\_2](https://doi.org/10.1007/1-4020-3933-6_2)
- Bickmore, T. W., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 12(2), 293–327.
- Blake, R. J. (2013). *Brave new digital classroom: Technology and foreign language learning* (2nd ed.). Georgetown University Press.
- Blake, R. J. (2017). Technologies for teaching and learning L2 speaking. In C. A. Chapelle & S. Sauro (Eds.), *The handbook of technology and second language teaching and learning* (pp. 107–117). Wiley-Blackwell.
- Braun, V., & Clarke, V. (2006). Qualitative research in psychology using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- Braun, V., & Clarke, V. (2012) Thematic analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol. 2: Research designs: Quantitative, qualitative, neuropsychological, and biological* (pp. 57–71). American Psychological Association.
- Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise, and Health*, 11(4), 589–597. <http://doi.org/10.1080/2159676X.2019.1628806>



- Brown, J. D. (2009). Open-response items in questionnaires. In J. Heigham & R. A. Croker (Eds.), *Qualitative research in applied linguistics* (pp. 200–219). Palgrave Macmillan.
- Bryant, C., Felice, M., Andersen, Ø. E., & Briscoe, T. (2019). The BEA-2019 shared task on grammatical error correction. In *Proceedings of the fourteenth workshop on innovative use of NLP for building educational applications*, 52–75, Italy. Association for Computational Linguistics. <https://www.aclweb.org/anthology/W19-4406>
- Burden, D., & Savin-Baden, M. (2019). *Virtual humans today and tomorrow*. CRC Press.
- Burns, A. (2016). Research and the teaching of speaking in the second language classroom. In E. Hinkel, (Ed.), *Handbook of research in second language teaching* (pp. 242–256). Routledge.
- Bygate, M. (2011). Teaching the spoken foreign language. In K. Knapp & B. Seidlhofer (Eds.), *Handbooks of applied linguistics communication competence language and communication problems practical solutions* (pp. 401–438). Walter de Gruyter.
- Call Annie. (2023). App Store Preview. *Call Annie: Video call your best AI friend*. Animato Inc. <https://apps.apple.com/us/app/call-annie/id6447928709>
- Casañ-Pitarch, R., & Gong, J. (2021). Testing ImmerseMe with Chinese students: Acquisition of foreign language forms and vocabulary in Spanish. *Language Learning in Higher Education*, 11(1), 219–233. <https://doi.org/10.1515/cercles-2021-2016>
- Cassell, J., & Bickmore, T. (2003). Negotiated collusion: Modeling social language and its relationship effects in intelligent agents. *User modeling and user-adapted interaction* 13, pp. 89–132. <https://doi.org/10.1023/A:1024026532471>

- Çelik, S., & Aytin, K. (2014). Teachers' views on digital educational tools in English language learning: Benefits and challenges in the Turkish context. *TESL-EJ*, 18(2), 1–18. <https://tesl-ej.org/wordpress/issues/volume18/ej70/ej70a1/>
- Cerratto Pargman, T., Lindberg, Y., & Buch, A. (2023). Automation is coming! Exploring future(s)-oriented methods in education. *Post-digital Science Education* 5, 171–194. <https://doi.org/10.1007/s42438-022-00349-6>
- Chapelle, C. A., & Sauro, S. (2017). Introduction. In C. A. Chapelle, & S. Sauro (Eds.), *The Handbook of technology and second language teaching and learning* (pp. 107–117). Wiley-Blackwell.
- Cheong Hin Hong, W. (2023). The impact of ChatGPT on foreign language teaching and learning: Opportunities in education and research. *Journal of Educational Technology and Innovation* (JETI) 5(1). <https://jeti.thewsu.org/index.php/cieti/article/view/103/64>
- Chilufy, E. M., & Arvola, M. (2021). Conceptual designing of a virtual receptionist: Remote desktop walkthrough and bodystorming in VR. In *Proceedings of the 9th International Conference on Human-Agent Interaction (HAI '21)*. Association for Computing Machinery, USA, (pp.112–120). <https://doi.org/10.1145/3472307.3484171>
- Chong, S. W., & Reinders, H. (2020). Technology-mediated task-based language teaching: A qualitative research synthesis. *Language Learning & Technology*, 24(3), 70–86. <http://hdl.handle.net/10125/44739>
- Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital games, design, and learning: A systematic review and meta-analysis. *Review of Educational Research*, 86(1), 79–122. <https://doi.org/10.3102/0034654315582065>
- Clarke, C. C. (1918). The phonograph in modern language teaching. *The Modern Language Journal*, 3(3), 116–122.

- Cleveland-Innes, M., Stenbom, S., & Hrastinski, S. (2014). *The influence of emotion on cognitive presence in a case of online math coaching*. [Paper presentation]. The 8th European Distance and E-Learning Network Research Workshop, UK.
- Collins, A., & Halverson, R. (2010). The second educational revolution: Rethinking education in the age of technology. *Journal of Computer Assisted Learning*, 26(1), 18–27. <https://doi.org/10.1111/j.1365-2729.2009.00339.x>
- Conti, G. (2022, May 18). The real MARSEARS: how an EPI sequence TRULY unfolds. *The Language Gym*. <https://gianfrancoconti.com/2022/05/18/the-real-marsears-how-an-epi-sequence-truly-unfolds>
- Conti, G., & Smith, S. (2016). *The language teacher toolkit*. Creatorspace independent publishing platform.
- Council of Europe. (2020). *Common European framework of reference for languages: Learning, teaching, assessment – Companion volume*, Council of Europe Publishing. [www.coe.int/lang-cefr](http://www.coe.int/lang-cefr).
- Council of Europe. (2022). *Enriching 21<sup>st</sup> century language education: The CEFR Companion volume in practice*. Council of Europe Publishing. <https://rm.coe.int/enriching-21st-century-language-education-the-cefr-companion-volume-in/1680a68ed0>
- Craig, S. D., & Schroeder, N. L. (2017). Reconsidering the voice effect when learning from a virtual human. *Computers & Education*, 114, 193–205. <https://doi.org/10.1016/j.compedu.2017.07.003>
- Craig, S. D., & Schroeder, N. L. (2018). Design principles for virtual humans in educational technology environments. In K. Millis, J. Magliano, D. Long, & K. Wiemer (Eds.) *Deep learning: Multi-disciplinary approaches* (pp. 128–139). Routledge/Taylor and Francis.

- Creswell, W. J., & Creswell, D. J. (2018). *Research design, qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage.
- Cucchiaroni, C., & Strik, H. (2019). Second language learners' spoken discourse: Practice and corrective feedback through automatic speech recognition. In M. Khosrow-Pour (Ed.), *Computer-assisted language learning: Concepts, methodologies, tools, and applications* (pp. 787–810). IGI Global.
- Darancik, Y. (2018). Students' views on language skills in foreign language teaching. *International Education Studies*, *11*(7), 166–178. <https://doi.org/10.5539/ies.v11n7p166>
- Dawson, S., Joksimovic, S., Mills, C., Gašević, D., & Siemens, G. (2023), Advancing theory in the age of artificial intelligence. *British Journal of Educational Technology*. <https://doi.org/10.1111/bjet.13343>
- DeKeyser, R. M. (2007). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*, *11*(2), 15th ed4. Cambridge University Press.
- de Vaus, D. (2001). *Research design in social research*. Sage Publications Ltd.
- DeVellis, R. F. (2007). *Scale development: theory and applications (applied social research)* (3rd Ed.). Sage Publications.
- Dewey, J. (1959). My pedagogic creed. In J. Dewey (Ed.), *Dewey on education* (pp. 19–32). Teachers College, Columbia University. (Original work published 1897)
- Dewey, J., & Bentley, A. F. (1999). Knowing and the known. In R. Handy & E. E. Hardwood (Eds). *Useful procedures of inquiry* (pp. 97–209). Behavioral Research Council. (First published in 1949)
- Dhimolea, T. K., Kaplan-Rakowski, R., & Lin, L. (2022). A systematic review of research on high-immersion virtual reality for language learning. *TechTrends*, *66*(5), 810–824. <https://doi.org/10.1007/s11528-022-00717-w>

- Dilmegani, C. (2023, March 20) *Generative AI, education. ChatGPT education use cases, benefits & challenges in 2023*. AI Multiple. <https://research.aimultiple.com/chatgpt-education/>
- Dimitriadis, Y. A. (2012). Supporting teachers in orchestrating CSCL classrooms. In A. Jimoyiannis (Ed.) *Research on e-Learning and ICT in Education*. Springer. [https://doi.org/10.1007/978-1-4614-1083-6\\_6](https://doi.org/10.1007/978-1-4614-1083-6_6)
- Divekar, R. R., Drozdal, J., Chabot, S., Zhou, Y., Su, H., Chen., Zhu, H., Hendler, J. A., & Braasch, J. (2021). Foreign language acquisition via artificial intelligence and extended reality: Design and evaluation. *Computer Assisted Language Learning*, 35(9), 2332–2360. <https://doi.org/10.1080/09588221.2021.1879162>
- D’Mello, S., Craig, S.D., Witherspoon, A., McDaniel, B., & Graesser, A. (2008). Automatic detection of learner’s affect from conversational cues. *User Modeling and User-Adapted Interaction*, 18(1–2), 45–80.
- Dreyfus, H. L., & Dreyfus, S. E. (1988). *Mind over machine: The power of human intuition and expertise in the era of the computer*. Free Press.
- Dunlosky, J., & Metcalfe, J. (2008). *Metacognition*. Sage Publications.
- Duolingo. (n.d.). *The free, fun, and effective way to learn a language!* <https://www.duolingo.com/>
- Dörnyei, Z., & Taguchi, T. (2010). *Questionnaires in second language research: Construction, administration, and processing* (2nd ed.). Routledge.
- Ebadi, S., & Amini, A. (2022). Examining the roles of social presence and human-likeness on Iranian EFL learners’ motivation using artificial intelligence technology: A case of CSIEC chatbot, *Interactive Learning Environments*, 1–19, <https://doi.org/10.1080/10494820.2022.2096638>

- Edholm, L. (2022). *Digitalisering i skolan har varit ett experiment*. [Digitalisation in school has been an experiment, my translation from Swedish]. Government Office of Sweden. <https://www.regeringen.se/debattartiklar/2022/12/digitaliseringen-i-skolan-har-varit-ett-experiment/>
- EdSurge. (2021). *Edtech investment around the world in 2020* [Infographic]. EdSurge Edtech Business. <https://www.edsurge.com/news/2021-03-11-edtech-investment-around-the-world-in-2020-infographic>
- Edison Research. (2021, March 11). *The Infinite dial 2021* [Online blog post]. <https://www.edisonresearch.com/the-infinite-dial-2021-2/>
- Egbert, J., Preuss, C., Huff, L., Sellen, J., & McNeil, L. (2009). Pedagogy, process, and classroom context: Integrating teacher voice and experience into research on technology-enhanced language learning. *The Modern Language Journal*, 93, 754–768. <http://www.jstor.org/stable/25612272>
- Ejlertsson, G. (2019). *Enkäten i Praktiken. En handbok i enkätmetodik*. [The survey in practice. A handbook in survey methodology, my translation from Swedish] (4<sup>th</sup> Ed.). Studentlitteratur.
- Ellis, R. (2005). Principles of instructed language learning. *System*, 33(2), 209–224.
- Ellis, R., Loewen, S., Elder, C., Erlam, R., Philip, J., & Reinders, H. (2009). *Implicit and explicit knowledge in second language learning, testing and teaching*. Multilingual Matters.
- Ellis, R., & Shintani, N. (2014). *Exploring language pedagogy through second language acquisition research*. Routledge.
- Ellis, R., Skehan, P., Li, A., Shintani, N., & Lambert, C. (2020). *Task-based language teaching: Theory and practice*. Cambridge University Press.

- ELSA. (2023). ELSA Home Page. <https://elsaspeak.net/>
- Erickson, G., Österberg, R., & Bardel, C. (2018) Lärares synpunkter på ämnet Moderna språk – en rapport från projektet TAL [Teachers' views on the subject modern languages – a report from the TAL project, my translation from Swedish], *LMS-Lingua*, 2, 8-12.
- Erickson, G., Bardel, C., Österberg, R., & Rosén, M. (2022). Attitudes and ambiguities: Teachers' views on second foreign language education in Swedish compulsory school. In C., Bardel, C., Hedman, K., Rejman, & E., Zetterholm (Eds.), *Exploring language education: Global and local perspectives* (pp. 157–223). Stockholm University Press. <https://doi.org/10.16993/bbz>
- Ethics Review Act [*Etikprövningslagen*]. SFS 2003:460. (2003). <https://lagen.nu/2003:460>
- European Commission. (2019a). *Key competencies for lifelong learning. Publications Office of the EU*. <https://op.europa.eu/en/publication-detail/-/publication/297a33c8-a1f3-11e9-9d01-01aa75ed71a1/language-en>
- European Commission. (2019b). *Education and training: Council recommendation on a comprehensive approach to the teaching and learning of languages*. [https://ec.europa.eu/education/education-in-the-eu/council-recommendation-improving-teaching-and-learning-languages\\_en](https://ec.europa.eu/education/education-in-the-eu/council-recommendation-improving-teaching-and-learning-languages_en)
- European Commission / EACEA / Eurydice. (2023). *Key data on teaching languages at school in Europe: 2023 edition*. Eurydice report. Publications Office of the European Union.
- European Parliament. (2017). *Language policy*. Fact Sheets on the European Union. [https://www.europarl.europa.eu/RegData/etudes/fiches\\_techniques/2013/051306/04A\\_FT\(2013\)051306\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/fiches_techniques/2013/051306/04A_FT(2013)051306_EN.pdf)

- Evans, S. (2023, April, 12). *ChatGPT makes humanoid robot multilingual: Engineered arts is using the technology to teach its robot new language skills*. AI Business. <https://aibusiness.com/verticals/chatgpt-gives-humanoid-robot-ability-to-speak-in-multiple-languages>
- Fernández, S. S., & Andersen, H. L. (2019). Oral proficiency in second and third foreign languages in the Danish education system. *Apples: Journal of Applied Language Studies*, 13, 49–67.
- Field, J. (2011). Cognitive validity. In L. Taylor (Ed.), *Examining speaking: Research and practice in assessing second language speaking* (pp. 65–111), Cambridge University Press.
- Field, A. (2018). *Discovering statistics using IBM SPSS statistics*. (5th ed.). Sage edge.
- Figueras, C., Verhagen, H., & Cerratto Pargman, T. (2022). Exploring tensions in Responsible AI in practice. An interview study on AI practices in and for Swedish public organizations. *Scandinavian Journal of Information Systems*. 34(2), 199–232. <https://aisel.aisnet.org/sjis/vol34/iss2/6>
- Flege, J. E. (2009). Give input a chance! In T. Piske, & M. Young-Scholten (Eds.), *Input matters in SLA* (pp. 175–190). Multilingual Matters.
- Franzke, A. S., Bechmann, A., Zimmer, M., & Ess, C. (2019). *Internet research: Ethical guidelines 3.0: Association of internet researchers*. <https://aoir.org/reports/ethics3.pdf>.
- Friedman, T. L. (2023, March 21). *Our Promethean moment*. The New York Times. <https://www.nytimes.com/2023/03/21/opinion/artificial-intelligence-chatgpt.html>
- Fryer, L. K., Nakao, K., & Thompson, A. (2019). Chatbot learning partners: Connecting learning experiences, interest and competence. *Computers in Human Behavior*, 93, 279–289. <https://doi.org/10.1016/j.chb.2018.12.023>



- Fryer, L. K., Coniam, D., Carpenter, R., & Lăpuşneanu, D. (2020). Bots for language learning now: Current and future directions. *Language Learning & Technology*, 24(2), 8–22. <http://hdl.handle.net/10125/44719>
- Gallacher, A., Thompson, A., & Howarth, M. (2018). “My robot is an idiot!” – Students’ perceptions of AI in the L2 classroom. In *Future-proof CALL: Language learning as exploration and encounters – short papers from EUROCALL 2018* (pp. 70–76). Research-publishing.net. <https://doi.org/10.14705/rpnet.2018.26.815>
- Garbati, J. F., & Mady, C. J. (2015). Oral skill development in second languages: A review in search of best practices. *Theory and Practice in Language Studies*, 5(9), 1763. <https://doi.org/10.17507/tpls.0509.01>
- Garrison, D. R., Anderson, T., & Archer, W. (2000). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education*, 2(2–3), 87–105.
- Gass, S., M. & Mackey, A. (2012). (Eds.). *The Routledge handbook of second language acquisition*. Routledge.
- Gass, S. M., & Mackey, A. (2015). Input, interaction, and output in second language acquisition. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (2nd ed., pp. 180–206). Routledge.
- Gibbons, P. (2015). *Teaching English language learners in the mainstream classroom*. Heinemann.
- GDPR. (2018). General Data Protection Regulation. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN>
- Godwin-Jones, R. (2021). Evolving technologies for language learning. *Language Learning & Technology*, 25(3), 6–26. <http://hdl.handle.net/10125/73443>

- Goh, C. C. M., & Burns, A. (2012). *Teaching speaking: A holistic approach*. Cambridge University Press.
- Goh, C. C. M. (2017). Research into practice: Scaffolding learning processes to improve speaking performance. *Language Teaching*, 50(2), 247–260. <https://doi.org/10.1017/S0261444816000483>
- Golonka, E.M., Bowles, A.R., Frank, V.M., Richardson, D.L., & Freynik, S. (2014). Technologies for foreign language learning: A review of technology types and their effectiveness. *Computer Assisted Language Learning*, 27, 70–105.
- González-Lloret, M. (2017). Technology for task-based language teaching. In C. A. Chapelle & S. Sauro (Eds), *The handbook of technology and second language teaching and learning* (pp. 234–247). Wiley-Blackwell.
- Google Assistant. (n.d.). *Hey Google. Make Google adjust it*. [https://assistant.google.com/intl/en\\_uk/](https://assistant.google.com/intl/en_uk/)
- Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American Psychologist*, 59(2), 93–104. <https://doi.org/10.1037/0003-066X.59.2.93>
- Government Office. (2017). Nationell digitaliseringsstrategi för skolväsende. [*National digital strategy for education*, my translation from Swedish] Utbildningsdepartementet. Retrieved 2018 from <https://www.regeringen.se/contentassets/72ff9b9845854d6c8689017999228e53/nationell-digitaliseringsstrategi-for-skolvasendet.pdf>
- Government Office. (2018). Nationell inriktning för artificiell intelligens. [National approach for digital intelligence]. Retrieved 2018 from [https://www.regeringen.se/49a828/contentassets/844d30fb0d594d1b9d96e2f5d57ed14b/2018ai\\_webb.pdf](https://www.regeringen.se/49a828/contentassets/844d30fb0d594d1b9d96e2f5d57ed14b/2018ai_webb.pdf)
- Government Office. (n.d.). *National minorities*. <https://www.regeringen.se/regeringens-politik/nationella-minoriteter/>

- Government Office (2022). Fortsatt strategi för skolans digitalisering [Continued strategy for the digitalisation of school, my translation from Swedish]. <https://www.regeringen.se/pressmeddelanden/2022/07/fortsatt-strategi-for-skolans-digitalisering>
- Graesser, A. C., Lippert, A. M., & Hampton, A. J. (2017). Successes and failures in building learning environments to promote deep learning: The value of conversational agents. In J. Buder & F. W. Hesse (Eds.), *Informational environments: Effects of use, effective designs* (pp. 273–298). Springer International Publishing [https://doi.org/10.1007/978-3-319-64274-1\\_12](https://doi.org/10.1007/978-3-319-64274-1_12)
- Granfeldt, J., Bardel, C., Erickson, G., Sayehli, S., Ågren, M., & Österberg, R. (2016). Muntlig språkfärdighet i främmande språk – en studie av samspelet mellan lärande, undervisning och bedömning i svensk skolkontext : presentation av det nya forskningsprojektet TAL [Oral proficiency in foreign languages – a study of the relation between learning, teaching, and assessment in a Swedish school context: presentation of the new research project TAL, my translation from Swedish]. *LMS - Lingua 2*, 1–6.
- Granfeldt, J. (2019). Muntlig språkfärdighet i främmande språk - en studie av samspelet mellan lärande, undervisning och bedömning [Oral proficiency in foreign languages – a study of the relation between learning, teaching, and assessment, my translation from Swedish], In *Resultatdialog 2019* (pp.29–32). VR Swedish Research Council. [https://www.vr.se/download/18.247403aa16e8e5ca6b3574/1574937055230/Resultatdialog\\_VR\\_2019.pdf](https://www.vr.se/download/18.247403aa16e8e5ca6b3574/1574937055230/Resultatdialog_VR_2019.pdf)
- Granfeldt, J., Sayehli, S., & Ågren, M. (2021). Trends in the study of modern languages in Swedish lower secondary school (2000 – 2018) and the impact of grade point average enhancement credits, *Education Inquiry*, 12(2), 127–146, <https://doi.org/10.1080/20004508.2020.1790099>

- Grant, S. J., Huang, H., & Pasfield-Neofitou, S. E., (2013). Language learning in virtual worlds: The role of foreign language and technical anxiety. *Journal For Virtual Worlds Research*, 6(1), 1–9. <https://journals.tdl.org/jvwr/index.php/jvwr/article/view/7027/6306>
- Grgurovic, M (2017). Blended Language learning: Research and practice. In C. A. Chapelle & S. Sauro (Eds.), *The handbook of technology and second language teaching and learning* (pp. 184–201). Wiley-Blackwell.
- Griol, D., & Callejas, Z. (2016). Conversational agents as learning facilitators: Experiences with a mobile multimodal dialogue system architecture. In S. Caballé, & R. Clarisó, (Eds.), *Formative assessment, learning data analytics and gamification in ICT Education* (pp 313-331). Academic Press
- Haelermans, C. (2017). *Digital tools in education: On usage, effects and the role of the teacher*. Studieförbundet Näringsliv och Samhälle SNS
- Harb, J., Abu Bakar, N., & Krish, P. (2014). Gender differences in attitudes towards learning oral skills using technology. *Education and Information Technologies*, 19(4), 805–816. <https://doi.org/10.1007/s10639-013-9253-0>
- Harlow, L. L., & Muyskens, J. A. (1994). Priorities for intermediate-level language instruction. *The Modern Language Journal*, 78, 141–154.
- Hartshorn, K. J., & McMurry, B., L. (2020). The effects of the COVID-19 pandemic on ESL learners and TESOL practitioners in the United States. *International Journal of TESOL Studies*, Vol. 2 (2), 140
- Hassani, K., Nahvi, A., & Ahmadi, A. (2016). Design and implementation of an intelligent virtual environment for improving speaking and listening skills. *Interactive Learning Environments*, 24(1), 252–271. <https://doi.org/10.1080/10494820.2013.846265>

- Heidari, S., Babor, T. F., De Castro, P., Tort, S., & Curno, M. (2016). Sex and gender equity in research: Rationale for the SAGER guidelines and recommended use. *Research Integrity and Peer Review* 1(2) <https://doi.org/10.1186/s41073-016-0007-6>
- Heift, T., & Chapelle, C. A. (2012). Language learning through technology. In S. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 555–569). Routledge
- Heigham, J. & Croker, R. A. (Eds.). (2009). *Qualitative research in applied linguistics*. Palgrave Macmillan. <http://dx.doi.org/10.1057/9780230239517>
- Hektner, J. M., Schmidt, J. A. & Csikszentmihalyi, M. (2007). *Experience sampling method -measuring the quality of everyday life*. Sage Publishing.
- Hodges, H., Steffensen, S. V., & Martin, J. E. (2012). Caring, conversing, and realizing values: New directions in language studies, *Language Sciences*, 34(5), 499–506. <https://doi.org/10.1016/j.langsci.2012.03.006>
- Howell, H., & Mikeska, J. N. (2021) Approximations of practice as a framework for understanding authenticity in simulations of teaching. *Journal of Research on Technology in Education*, 53(1), 8–20, <https://doi.org/10.1080/15391523.2020.1809033>
- Hrastinski, S., Olofsson, A. D., Arkenback, C., Ekström, S., Ericsson, E., Fransson, G., ... Utterberg, M. (2019). Critical imaginaries and reflections on artificial intelligence and robots in postdigital K-12 education. *Postdigital Science Education* 1, 427–445. <https://doi.org/10.1007/s42438-019-00046-x>
- Hsu, M. M., Chen, P. S., & Yu, C. S. (2021). Proposing a task-oriented chatbot system for EFL learners' speaking practice, *Interactive Learning Environments*, 1–12. <https://doi.org/10.1080/10494820.2021.1960864>

- Huang, W., Foon Hew, K. & Fryer, L. K. (2021). Chatbots for language learning: Are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning* 38(1), 237–257. <https://doi.org/10.1111/jcal.1261>
- Huang, Y. C., Backman, S. J., Backman, K. F., McGuire, F. A., & Moore, D. W. (2019). An investigation of motivation and experience in virtual learning environments: A self-determination theory. *Education and Information Technologies*, 24(1), 591–611. <https://doi.org/10.1007/s10639-018-9784-5>
- Huang, Y.-T., Lin, Y.-L., Yang, J.-C., & Wu, Y.-C. (2008). *An English dialogue companion system for supporting conversation practice*. Proceedings of the 16th International Conference on Computers in Education ICCE 2008.
- Humble, N., & Mozelius, P. (2022). The threat, hype, and promise of artificial intelligence in education. *Discover Artificial Intelligence*, 2(1). <https://doi.org/10.1007/s44163-022-00039-z>
- Hwang, G. J., & Chang, C. Y. (2021). A review of opportunities and challenges of chatbots in education. *Interactive Learning Environments*, 1–14. <https://doi.org/10.1080/10494820.2021.1952615>
- Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics: Selected readings* (pp. 269–293). Penguin Books.
- ICJME (2023). International Committee of Medical Journal Editors. *Defining the Role of Authors and Contributors*. <https://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html>
- ImmerseMe. (n.d.). *VR-based language learning*. <https://immerseme.col>
- Jarvis, P. (2006). *Towards a comprehensive theory of human learning*. Routledge.

- Jeon, J. (2022). Exploring AI chatbot affordances in the EFL classroom: Young learners' experiences and perspectives. *Computer Assisted Language Learning, Ahead-of-print(Ahead-of-print)*, 1–26. <https://doi.org/10.1080/09588221.2021.2021241>
- Ji, H., Han, I., & Ko, Y. (2022). A systematic review of conversational AI in language education: Focusing on the collaboration with human teachers. *Journal of Research on Technology in Education*, 55(1), 48–63, Routledge. <https://doi.org/10.1080/15391523.2022.2142873>
- Johnson, K. (1996). *Language teaching and skill learning*. Blackwell
- Johnson, W. L. & Wang, N. (2008). The politeness effect in an intelligent foreign language tutoring system. In B. P. Woolf, E. Aimeur, R. Nkambou, & S. Lajoie (Eds.), *Intelligent tutoring systems Proceedings of the ITS 2008* (pp. 270–280). Springer. [http://doi.org/10.1007/978-3-540-69132-7\\_31](http://doi.org/10.1007/978-3-540-69132-7_31)
- Johnson, W. L., & Valente, A. (2009). Tactical language and culture training systems: Using AI to teach foreign languages and cultures. *AI Magazine*, 30(2), 72–83. <https://doi.org/10.1609/aimag.v30i2.2240>
- Johnson, W. L., & Lester, J. C. (2016). Face-to-face interaction with pedagogical agents, twenty years later. *International Journal of Artificial Intelligence in Education*, 26, 25–36. <https://doi.org/10.1007/s40593-015-0065-9>
- Johnson, W. L. (2019a). Data-driven development and evaluation of Enskill English. *International Journal of Artificial Intelligence in Education*, 29, 425–457. <http://doi.org/10.1007/s40593-019-00182-2>.
- Johnson, L. (2019b). Using AI simulations to build fluency and confidence. Alelo Inc. [White paper] <https://www.alelo.com/wp-content/uploads/2022/02/Enskill-White-Paper-Using-AI-Simulations-to-Build-Fluency-190708.pdf>

- Johnson, L. (2020). Case study: UVM program using Alelo Enskill increases students' spoken English proficiency. Alelo Inc. [White paper] [https://www.alelo.com/wp-content/uploads/2022/02/White Paper UVM Case Study 200113-1.pdf](https://www.alelo.com/wp-content/uploads/2022/02/White-Paper-UVM-Case-Study-200113-1.pdf)
- Johnson, W. L. (2021). Case study: Rapid training of community health workers using Enskill®. *International Journal of Advanced Corporate Learning (IJAC)*, 14(2), 44–49. <https://doi.org/10.3991/ijac.v14i2.25929>
- Kaplan-Rakowski, R., & Gruber, A. (2019). Low-immersion versus high-immersion virtual reality: Definitions, classification, and examples with a foreign language focus. In *Proceedings of the Innovation in Language Learning International Conference 2019*. Pixel.
- Kaplan-Rakowski, R., & Wojdyski, T. (2018). Students' attitudes toward high-immersion virtual reality assisted language learning. In P. Taalas, J. Jalkanen, L. Bradley, & S. Thouëсны (Eds.), *Future-proof CALL: Language learning as exploration and encounters – short papers from EUROCALL 2018* (pp. 124–129). <https://doi.org/10.14705/rpnet.2018.26.824>
- Kassim, S., Witkin, N., & Stone, A. (2019). Student perceptions of virtual reality use in a speaking activity. In F. Meunier, J. V. de Vyver, L. Bradley, & S. Thouëсны (Eds.), *CALL and complexity – Short papers from EUROCALL 2019* (pp. 223–228). Research-publishing.net. <https://doi.org/10.14705/rpnet.2019.38.1013>
- Kim, H.-S, Cha, Y., & Kim, N. Y. (2021). Effects of AI chatbots on EFL students' communication skills. *Korean Journal of English Language and Linguistics*, 21, 712–734 <https://doi.org/10.15738/kjell.21..202108.712>
- Kim, N.-Y. (2016). Effects of voice chat on EFL learners' speaking ability according to proficiency levels. *Multimedia-Assisted Language Learning*, 19(4), 63–88. <https://doi.org/10.15702/mall.2016.19.4.63>



- Knapp, K., Seidlhofer, B., & Widdowson, H. (2011). Introduction. Developing foreign language communication: Principles and practices. In K. Knapp, B. Seidlhofer, & H. Widdowson (Eds). *Handbook of foreign language communication and learning* (pp. 1–20). De Gruyter Mouton.
- Kocaballi, A. B., Laranjo, L. & Coiera, E. (2018). Measuring user experience in conversational interfaces: A comparison of six questionnaires. In *Proceedings of the 32nd International BCS Human Computer Interaction Conference (HCI)*. <http://doi.org/10.14236/ewic/HCI2018.21>
- Krashen, S. (1985). *The input hypothesis: Issues and implications*. Longman.
- Krumsvik, R. J., Berrum, E., & Jones, L. (2018). Everyday digital schooling: Implementing tablets in Norwegian primary school. Examining outcome measures in the 1st cohort. *Nordic Journal of Digital Literacy* 3(13), 152–178. <https://www.idunn.no/dk/2018/03/everyday-digital-schooling-implementing-tablets-in-norweg>
- Kukulska-Hulme, A., & Lee, H. (2020). Intelligent assistants in language learning: An analysis of features and limitations. In *CALL for widening participation: Short papers from EUROCALL 2020* (pp. 172–176). Research-publishing.net. <https://doi.org/10.14705/rpnet.2020.48.1184>
- Källström, Å., & Andersson Bruck, K. (2017). *Etiska reflektioner i forskning med barn [Ethical reflections in research with children, my translation from Swedish]*. Gleerups.
- Lackéus, M. (2020). Collecting digital research data through social media platforms: Can ‘scientific ‘social media’ disrupt entrepreneurship research methods? In W. B. Gartner & B. Teague (Eds.), *Research handbook of entrepreneurial behavior, practice, and process* (pp.199–241). Edward Elgar Publishing

- Lackéus, M. (2021). *Den vetenskapande läraren – en handbok för forskning i skola och förskola [The scientific teacher – a handbook for research in schools and preschools, my translation from Swedish]*. Studentlitteratur.
- Lantolf, J., Thorne, S. L., & Poehner, M. E. (2014). Sociocultural theory and second language development. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 207–226). Taylor and Francis.
- Lantolf, J. P., & Swain, M. (2020). Perezhivanie: The cognitive-emotional dialectic within the social situation of development. In A. H. Al-Hoorie & P. D. MacIntyre (Eds.), *Contemporary language motivation theory: 60 years since Gardner and Lambert (1959)* (pp. 80–105). Multilingual Matters.
- Lawson, A. P., Mayer, R. E., Adamo-Villani, N., Benes, B., Lei, X., & Cheng, J. (2021). Do learners recognize and relate to the emotions displayed by virtual instructors? *International Journal of Artificial Intelligence in Education*, 31(1), 134–153. <https://doi.org/10.1007/s40593-021-00238-2>
- Lee, J. H., Kim, J., & Kim, Y. (2018) Effects of conversational agents on EFL learners' oral production and motivation. *System*, 78, 150–162.
- Letmark, P. (2023, May, 14). Regeringen river upp Skolverkets plan för digitalisering i skolan. [*The Government tears up the new plan for digitalisation of school, my translation from Swedish*] Dagens Nyheter. <https://www.dn.se/sverige/regeringen-river-upp-skolverkets-plan-for-digitalisering-i-skolan/>
- Leviathan, Y. & Matias, Y. (2018, May 8). *Google duplex: An AI system for accomplishing real-world tasks over the phone*. Google Research. <https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html>
- Levy, M. (1990). Towards a theory of CALL. *CALL Journal*, 1(4), 5–7.

- Levy, M. (1997). *Computer-assisted language learning: Context and conceptualization*. Clarendon Press.
- Levy, M. (2015). The role of qualitative approaches to research in CALL contexts: Closing in on the learner's experience. *CALICO Journal*, 32(3), 554–568. <https://doi.org/10.1558/cj.v32i3.2660>
- Li, L. (2017). *New technologies and language learning*. Palgrave.
- Lin, T. J., & Lan, Y. J. (2015). Language learning in virtual reality environments: Past, present, and future. *Educational Technology & Society*, 18 (4), 486–497.
- Littlewood, W. (1984). *Communicative language teaching: An introduction*. Cambridge University Press.
- Littlewood, W. (2004). The tasked-based approach: Some questions and suggestions. *ELT Journal*, 158(4), 319–326.
- Littlewood, W. (2018). Developing a personal approach to teaching language for communication. *Journal of Asia TEFL*, 15(4), 1222–1229.
- Liu, C., & Forsythe, S. (2021). The effect of virtual assistant anthropomorphic features on customer satisfaction and loyalty. *International Journal of Hospitality Management*, 94, 102863. <https://doi.org/10.1016/j.ijhm.2021.102863>
- Loewen, S. (2012). The role of feedback. In S. M. Gass, & A. Mackey, (Eds.), *The Routledge handbook of second language acquisition* (pp. 24–40). Taylor and Francis.
- Loewen, S., & Sato, M. (2018). Interaction and instructed second language acquisition. *Language Teaching*, 51(3), 285–329. <https://doi.org/10.1017/S0261444818000125>

- Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. C Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). Academic Press.
- Long, M. (2015). *Second language acquisition and task-based language teaching*. Wiley-Blackwell.
- LoopMe. (2023). LoopMe. *LoopMe for researchers and evaluators; How to collect trustworthy and relevant data?* <https://loopme.se/solutions/research-and-evaluation/>
- MacIntyre, P. D., Dörnyei, Z., Clément, R., & Noels, K. A. (1998). Conceptualizing willingness to communicate in an L2: A situational model of L2 confidence and affiliation. *The Modern Language Journal*, 82(4), 545–562. <https://doi.org/10.1111/j.1540-4781.1998.tb05543.x>
- Mahn, H., & Fazalehaq, H. M. (2020). Vygotsky and second language acquisition. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics*. <https://doi.org/10.1002/9781405198431.wbeal1272.pub2>
- Mariani, L. (1997). Teacher support and teacher challenge in promoting learner autonomy. *Perspectives: A Journal of TESOL Italy*, 23(2), 5-19.
- Maring, J. (2023, April 27). *Call Anne* [Photo on blog post]. Digital Trends. <https://www.digitaltrends.com/wp-content/uploads/2023/04/call-annie-chatgpt-facetime-app.jpg?fit=720%2C479&p=1>
- Markham, A., & Buchanan, E. (2012). *Ethical decision-making and Internet research: Recommendations from the AoIR ethics working committee (version 2.0)*. <http://aoir.org/reports/ethics2.pdf>
- McNamara, D., Allen, L., Crossley, S., Dascalu, M., & Perret, C. (2017) Natural language processing and learning analytics. In C. Lang, G. Siemens, A. Wise, & D. Gasevic, D. (Eds.), *Handbook of learning analytics* (pp. 93-104). SOLAR, Society for Learning Analytics and Research.

- McTear, M. (2020). *Conversational AI: Dialogue systems, conversational agents, and chatbots*. Morgan & Claypool.
- Meri-Yilan, S. (2019). A constructivist desktop virtual reality-based approach to learning in a higher education institution. In K. Becnel (Ed.), *Emerging Technologies in Virtual Learning Environments*, (pp. 258–283). IGI Global. <https://doi.org/10.4018/978-1-5225-7987-8.ch013>
- Miller, J., & Glassner, B. (2016). The “inside” and the “outside”: Finding realities in interviews. In D. Silverman. *Qualitative research* (Ed.), (pp. 51–66). Sage.
- Mills C., & D’Mello, S. (2014). On the validity of the autobiographical emotional memory task for emotion induction. *PLoS ONE* 9(4). <https://doi.org/10.1371/journal.pone.0095837>
- Mitchell, R., Myles, F., & Marsden, E. (2012). *Second language theories*. (3rd ed.). Routledge Taylor & Francis Group.
- Mondly. (2023). Pearson. *Play your way to a new language*. <https://www.mondly.com/>
- Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, 19(2), 98–100, <http://doi.org/10.1109/MRA.2012.2192811>
- Morton, H., & Jack, M. A. (2005). Scenario-based spoken interaction with virtual agents. *Computer Assisted Language Learning*, 18(3), 171–191. <https://doi.org/10.1080/09588220500173344>
- Morton, H., Gunson, N., & Jack, M.A. (2012). Interactive language learning through speech enabled virtual scenarios. *Advances in Human-Computer Interaction*, 2012, 1–14. <https://doi.org/10.1155/2012/389523>

- Moussalli, S., & Cardoso, W. (2020). Intelligent personal assistants: Can they understand and be understood by accented L2 learners? *Computer Assisted Language Learning*, 33(8), 865–890. <https://doi.org/10.1080/09588221.2019.1595664>
- Nation, P. (2007). The four strands, innovation. *Language Learning and Teaching*, 1(1), 2–13. <https://doi.org/10.2167/illt039.0>
- National Agency for Education. (2018a). *Curriculum for the compulsory school, preschool class and school-age educare*. <https://www.skolverket.se/getFile?file=3984>
- National Agency for Education. (2018b). *Digitaliseringen i skolan - möjligheter och utmaningar*. [Digitalisation in school: opportunities and challenges, my translation from Swedish] <https://www.skolverket.se/download/18.6bfaca41169863e6a65d438/1553968018488/pdf3971.pdf>
- National Agency for Education. (2020). *Elever och skolenheter i grundskolan läsåret 2019/20. Sveriges officiella statistik*. [Students and school units in primary school in the academic year 2019/20. Sweden's official statistics, my translation from Swedish]. Dnr 5.1.1-2019:387. <https://www.skolverket.se/getFile?file=6477>
- National Agency for Education. (2022a). *Förslag till nationell digitaliseringsstrategi för skolväsendet 2023–2027*. [Proposition of national digital strategy of education, my translation into English]. <https://www.skolverket.se/getFile?file=10849>
- National Agency for Education. (2022b). *Läroplan för grundskolan, förskoleklassen och fritidshemmet Lgr22*. [Curriculum for the compulsory school, preschool class, and the leisure-time centre] <https://www.skolverket.se/getFile?file=9718>
- National Agency for Education. (2022c) *Syllabus, English*. <https://www.skolverket.se/undervisning/grundskolan/laroplan-och-kursplaner-for-grundskolan/laroplan-lgr22-for-grundskolan-samt-for-for-skoleklassen-och-fritidshemmet?url=-996270488%2Fcompulsorycw%2Fjsp%2Fsubject.htm%3FsubjectCode%3DGRGRENG01%26tos%3Dgr&sv.url=12.5dfce44715d35a5cdfa219f>

- National Agency for Education. (2022d). *Syllabus, modern languages*. <https://www.skolverket.se/undervisning/grundskolan/laroplan-och-kursplaner-for-grundskolan/laroplan-lgr22-for-grundskolan-samt-for-forskoleklassen-och-fritidshemmet?url=-996270488%2Fcompulsorycw%2Fjsp%2Fsubject.htm%3FsubjectCode%3DGRGRMSP01%26tos%3Dgr&sv.url=12.5dfce44715d35a5cdfa219f>
- Nilsson, T. (2017.). “I keep returning to it, so something must be happening”: Swedish foreign language teacher students’ conceptions of ICT and digital tools in foreign language teaching and learning. *Innovation in Language Learning, International Conference*. <https://conference.pixel-online.net/ICT4LL/files/ict4ll/ed0011/FP/1337-LTT3414-FP-ICT4LL11.pdf>
- Nilsson, T. (2018). *Det språkliga inflödets betydelse för muntlig språkutveckling*. [The importance of input for development of speaking, my translation from Swedish]. Skolverket. <https://larportalen.skolverket.se>
- Nowell, L. S., Norris, J. M., White, D. E. & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods*, 16, 1–13. <https://doi.org/10.1177/1609406917733847>
- Ohta, A. S. (1995). Applying sociocultural theory to an analysis of learner discourse: Learner-learner collaborative interaction in the zone of proximal development. *Issues in Applied Linguistics*, 62, 93–121.
- OpenAI. (2022, November 30). ChatGPT: Optimizing language models. [blogpost] <https://openai.com/blog/chatgpt/>
- Otto, S. E. K. (2017). From past to present: A hundred years of technology for L2 learning. In C. A., Chapelle & S. Sauro (Eds.), *The Handbook of technology and second language teaching and learning* (pp. 10-25). Wiley-Blackwell.

- Pakula, H. (2019). Teaching speaking. *Apples – Journal of Applied Language Studies*, 13(1), 95–111. <https://doi.org/10.17011/apples/urn.201903011691>
- Pallant, J. (2011). *SPSS survival manual: A step by step guide to data analysing using SPSS*. (4th ed.). Allen & Unwin.
- Papert, S. (1993). *Mindstorms: Children, computers and powerful ideas*. Basic Books
- Papin, K. (2018). Willingness to communicate: Can virtual reality help? In P. Taalas, J. Jalkanen, L. Bradley, & S. Thouäсны (Eds.), *Future-proof CALL: Language learning as exploration and encounters: Short papers from EUROCALL 2018* (pp. 243-248). Research-publishing.
- Pekruna, R. (2020). Commentary: Self-report is indispensable to assess students' learning. *Frontline Learning Research*, 8(4), 185–193. European Association for Research on Learning and Instruction. <https://doi.org/10.14786/flr.v8i3.637>
- Pieraccini, R. (2012) *The voice in the machine: Building computers that understand speech*. MIT Press.
- Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press.
- Richards, J., & Rodgers, T. (2001). *Approaches and methods in language teaching*. Cambridge University Press.
- Robinson, K. E., Allen, P. J., Quail, M., & Beilby, J. (2020). Virtual patient clinical placements improve student communication competence. *Interactive Learning Environments*, 28(6), 795–805. <https://doi.org/10.1080/10494820.2018.1552869>



- Rosell-Aguilar, F. (2017). State of the app: A taxonomy and framework for evaluating language learning mobile applications. *CALICO Journal*, 34(2), 243–258. <https://doi.org/10.1558/cj.27623>
- Rosetta Stone. (2023). *The Experts in language learning*. <https://www.rosetta-stone.com/>
- Roth, W. M., & Jornet, A. (2014). Towards a theory of experience. In *Science Education*, 98(1), 106–126. <https://doi.org/10.1002/sci.21085>
- Sang, Y., & Hiver, P. (2021). Engagement and companion constructs in language learning: Conceptualising learners' involvement in the L2 classroom. In P. Hiver, A.H., Al-Hoorie & S. Mercer. (Eds.), *Student engagement in the language classroom* (pp. 17–37). Multilingual Matters.
- Sarwar, N. (2023, April 27). *You can now video chat with a ChatGPT AI: Here's what it looks like*. Digital Trends, [Blogpost]. <https://www.digitaltrends.com/mobile/chatgpt-ai-video-calling-app-call-annie-news/>
- Schroeder, N., Adesope, O., & Gilbert, R. (2013). How effective are pedagogical agents for learning? A meta-analytic review. *Journal of Educational Computing Research*, 49(1), 1–39. <https://doi.org/10.2190/EC.49.1.a>
- Schroeder, N. L., Chiou, E. K., & Craig, S. D. (2021). Trust influences perceptions of virtual humans, but not necessarily learning. *Computers & Education*, 160, 1–11. <https://doi.org/10.1016/j.compedu.2020.104039>
- Schroeder, N. L., & Craig, S. D. (2021). Learning with virtual humans: Introduction to the special issue. *Journal of Research on Technology in Education*, 53(1), 1–7. <https://doi.org/10.1080/15391523.2020.1863114>

- Schwienhorst, K. (2002). Why virtual, why environments? Implementing virtual reality concepts in computer-assisted language learning. *Simulation & Gaming*, 33(2), 196–209. <https://doi.org/10.1177/1046878102332008>
- Selwyn, N. (2023a). Constructive criticism? Working with (rather than against) the AIED back-lash. *International Journal of Artificial Intelligence in Education* [in press].
- Selwyn, N. (2023b, May 22). *TIMDA bjuder in till reflektion*. TIMDA Text, Information och multimodalitet i den digitala eran. [Blogpost about information and multimodality]. <https://timda.home.blog/2023/05/22/timda-bjuder-in-till-reflektion/>
- Shadish, W. R., Cook, T. D. & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin Company.
- Shenton, A. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information*, 22(2), 63–75.
- Sieloff Magnan, S., & Lafford, B. A. (2012). Learning through immersion during study abroad. In S. M. Gass, & A. Mackey (Eds.), *The Routledge Handbook of Second Language Acquisition* (pp. 525–540). Taylor and Francis.
- Skinner, B. F. (1968). *The technology of teaching*. Appleton-Century-Crofts.
- Skoldigiplan (2018). *Nationell handlingsplan för skolväsendet*. [National action plan for education, my translation from Swedish]. <http://www.skoldigiplan.se/>
- Sofkova Hashemi, S., & Cederlund, K. (2017). Making room for the transformation of literacy instruction in the digital classroom. *Journal of Early Childhood Literacy*, 17(2), 221–253. <https://doi.org/10.1177/1468798416630779>

- Sofkova Hashemi, S., Lindstrom, N. B., Bartram, L., & Bradley, L. (2017). Investigating mobile technology resources for integration: The Technology-Pedagogy-Language-Culture (TPLC) Model. In *ACM International Conference Proceeding Series*. Association for Computing Machinery. 20, 1-8. <https://doi.org/10.1145/3136907.3136930>
- Soto, J. H. B., Ocampo, D. C. T., del Carmen Beltrán Colón, L., & Oropesa, A. V. (2020). Perceptions of ImmerseMe virtual reality platform to improve English communicative skills in higher education. *International Journal of Interactive Mobile Technologies*, 14(7), 4–19. <https://doi.org/10.3991/IJIM.V14I07.12181>
- Stenbom, S. (2015). *Online coaching as a relationship of inquiry: Exploring one-to-one online education*. [Doctoral thesis in education and communication in the educational sciences, KTH, Stockholm].
- Stenbom, S., Jansson, M., & Hulkko, A. (2016). Revising the Community of Inquiry framework for the analysis of one-to-one online learning relationships. *The International Review of Research in Open and Distributed Learning*, 17(3), 37–53.
- Sundqvist, P., & Sylvé, L. (2014). Language-related computer use: Focus on young L2 English learners in Sweden. *ReCALL*, 26(1), 3–20. <https://doi.org/10.1017/S0958344013000232>
- Swain, M. (2013). The inseparability of cognition and emotion in second language learning. *Language Teaching*, 46(2), 195–207. <https://doi.org/10.1017/S0261444811000486>
- Swedish Ethics Review Authority (*Etikprövningsmyndigheten*). (n.d.). Värnar människan i forskning. <https://etikprovningmyndigheten.se/>
- Swedish Internet Foundation. (2019). *The Swedes and the Internet*, Chapter 1. Fiber ökar och äldre alltmer digitala. <https://svenskarnaochinternet.se/rapporter/svenskarna-och-internet-2019/allmant-om-internetutvecklingen/1-av-20-har-en-uppkopplad-hogtalare/>

- Swedish Research Council. (2017). *Good research practice*. <https://www.vr.se/english/analysis-and-assignments/we-analyse-and-evaluate/all-publications/publications/2017-08-31-good-research-practice.html>
- Swedish Research Council. (2019). *Research overview 2019: Educational research*. <https://www.vr.se/english/analysis/reports/our-reports/2019-05-10-research-overview-2019-educational-research.html>
- Swedish School Inspectorate. (2022). *Skolans arbete för kvalitet i undervisningen moderna språk i – Undervisning för kommunikativ förmåga [Quality report, modern languages]*
- Sydorenko, T., Daurio, P., & Thorne, S. L. (2018). Refining pragmatically-appropriate oral communication via computer-simulated conversations. *Computer Assisted Language Learning*, 31(1–2), 157–180. <http://doi.org/10.1080/09588221.2017.1394326>
- Säljö, R. (2022). *Lärande - en introduktion till perspektiv och metaforer [Learning: An introduction to perspectives and metaphors, my translation from Swedish]*. Gleerups Utbildning AB.
- Tai, T. Y., & Chen, H. H. J. (2022). The impact of intelligent personal assistants on adolescent EFL learners' speaking proficiency. *Computer Assisted Language Learning, Ahead-of-print*(Ahead-of-print), 1-28. <https://doi.org/10.1080/09588221.2022.2070219>
- TAL. (n.d.). *TAL-projektet. Learning, teaching and assessment of second foreign languages: An alignment study on oral language proficiency in the Swedish school Context*. Lund University. <https://www.tal.lu.se/projektbeskrivning/>
- Tallvid, M. (2015). *1:1 I klassrummet analyser av en pedagogisk praktik i förändring*. [Doctoral Thesis 42, Centre for Educational sciences and teacher research. Graduate School CUL, The University of Gothenburg].

- Tegmark, M. (2017). *Life 3.0: Being human in the age of artificial intelligence*. Knopf.
- Tewari, G. (2022). *Recent advancements in artificial intelligence*. Forbes. <https://www.forbes.com/sites/forbesbusinesscouncil/2022/10/07/recent-advancements-in-artificial-intelligence/>
- The University of Gothenburg. (2020). *Management of research data*. <https://medarbetarportalen.gu.se/service-support/management-of-research-data/>
- Thiriau, C. (2017). *Teaching speaking in ELT: Global teaching speaking questionnaire*. Cambridge University Press. <https://www.cambridge.org/elt/blog/wp-content/uploads/2017/11/Cambridge-global-teaching-speaking-questionnaire-2017.pdf>
- Tholin, J. (2019). State control and governance of schooling and their effects on French, German, and Spanish learning in Swedish compulsory school, 1996–2011. *Scandinavian Journal of Educational Research*, 63(3), 317–332. <http://doi.org/10.1080/00313831.2017.1375004>
- Thorell, L., Klingberg, T., & Hemlin, A. (2023, April 24). *Skolverkets förslag ett riskabelt experiment*. [The proposition of the Swedish National Agency for Education, a risky experiment, my translation from Swedish]. Svenska Dagbladet. <https://www.svd.se/a/764VzB/skolverkets-forslag-ett-riskabelt-experiment-skriver-thorell-klingberg-och-hemlin>
- Timpe-Laughlin, V., Sydorenko, T., & Daurio, P. (2020). Using spoken dialogue technology for L2 speaking practice: What do teachers think? *Computer Assisted Language Learning*, 35(5–6), 1194–1217. <https://doi.org/10.1080/09588221.2020.1774904>
- Timpe-Laughlin, V., Sydorenko, T., & Dombi, J. (2022). Human versus machine: Investigating L2 learner output in face-to-face versus fully automated role-plays. *Computer Assisted Language Learning*. <https://doi.org/10.1080/09588221.2022.2032184>

- Tran, M., N. (2021). *The use of virtual reality technology in foreign language education : From teachers' perspectives*. [Dissertation submitted in partial completion of the MSc in Education Applied Linguistics and Second Language Acquisition, Faculty of Education – The University of Oxford].
- Turing, A. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433–460. Oxford University Press. <http://www.jstor.org/stable/2251299>
- Ushioda, E., & Dörnyei, Z. (2012). Motivation. In S. M. Gass, & A. Mackey, (Eds.), *The Routledge handbook of second language acquisition* (pp. 396–409). Taylor and Francis.
- Utami, L. P. R. A., Suwastini, N. K. A., Dantes, G. R., Suprihatin, C. T., & Adnyani, K. E. K. (2021). Virtual reality for supporting authentic learning in 21st century language classroom. *Jurnal Pendidikan Teknologi Dan Kejuruan*, 18(1), 132. <https://doi.org/10.23887/jptk-undiksha.v18i1.32376>
- Utterberg Modén, M., Lundin, J., & Tallvid, M. (2020). *Artificiell intelligens i skolan kräver ökad insikt hos lärarna* [Artificial intelligence and the need for the insights of teachers, my translation from Swedish]. National Agency for Education. <https://www.skolverket.se/skolutveckling/forskning-och-utvarderingar/artiklar-om-forskning/artificiell-intelligens-i-skolan-kraver-okad-insikt-hos-lararna>
- Utterberg Modén, M., Lundin, J., & Tallvid, M. (2023). *Digitala läromedel i en digitaliserad skola* [Digital teaching materials in a digitalized school, my translation from Swedish]. Gleerups.
- van Doremalen, J., Boves, L., Colpaert, J., Cucchiari, C., & Strik, H. (2016). Evaluating automatic speech recognition-based language learning systems: A case study. *Computer Assisted Language Learning*, 29(4), 833–851. <https://doi.org/10.1080/09588221.2016.1167090>

- VanPatten, B., & Benati, A. G. (2015). *Key terms in second language acquisition* (2nd ed.). Bloomsbury.
- Van Pinxteren, M. M. E., Pluymaekers, M., & Lemmink, J. G. A. M. (2020). Human-like communication in conversational agents: A literature review and research agenda. *Journal of Service Management*, *31*(2), 203–225. <https://doi.org/10.1108/JOSM-06-2019-0175>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Walker, E., & Ogan, A. (2016). We're in this together: Intentional design of social relationships with AIED systems. *International Journal of Artificial Intelligence in Education* *26*, 713–729. <https://doi.org/10.1007/s40593-016-0100-5>
- Walker, A., & White, G. (2017). *Technology enhanced language learning*. Oxford University Press
- Walker, T. (2023). Chatbots.org. *161 Humanlike conversational AI synonyms*. <https://chatbots.org/synonyms>
- Wang, Y. F., Petrina, S., & Feng, F. (2017). VILLAGE–Virtual immersive language learning and gaming environment: Immersion and presence. *British Journal of Educational Technology*, *48*(2), 431–450. <https://doi.org/10.1111/bjet.12388>
- Wang, Q. (2020). The role of classroom-situated game-based language learning in promoting students' communicative competence. *International Journal of Computer-Assisted Language Learning and Teaching*, *10*(2), 59–82
- Wang, X., Pang, H., Wallace, M. P., Wang, Q., & Chen, W. (2022). Learners' perceived AI presences in AI-supported language learning: A study of AI as a humanized agent from community of inquiry. *Computer Assisted Language Learning*. <https://doi.org/10.1080/09588221.2022.2056203>

- Warschauer, M. & Kern, R. (2000). *Networked-based language teaching concepts and practice*. Cambridge University Press.
- Warschauer, M. (2004). Technological change and the future of CALL. In S. Fotos. & C. Brown (Eds.), *New perspectives on CALL for second and foreign language classrooms* (pp.15–25). Lawrence Erlbaum Associates.
- Weilenmann, A. (2021). Shared screen time: The role of the mobile phone in local social interaction in 2000 and 2020. In J. Katz, J. Floyd, & K. Schiepers (Eds.), *Perceiving the future through new communication technologies*. Palgrave Macmillan [https://doi.org/10.1007/978-3-030-84883-5\\_4](https://doi.org/10.1007/978-3-030-84883-5_4)
- Weizenbaum, J. (1996). ELIZA: A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36–45. <https://doi.org/10.1145/365153.365168>
- Wiliam, D. (2017). *Embedded formative assessment: Strategies for classroom assessment that drives student engagement and learning*. Solution Tree Press.
- Williamson, B. (2016). Boundary brokers: Mobile policy networks, database pedagogies, and algorithmic governance in education. In T. Ryberg, C. Sinclair, S. Bayne, & M. de Laat (Eds.), *Research, boundaries, and policy in networked learning* (pp. 41–57). Springer International Publishing.
- Winograd, T. (1972). *Understanding natural language*. Dissertation. Academic Press.
- Winograd, T. (1984). What does it mean to understand a language? *Cognitive Science* 4(3), 209–241.



- Wood, D. J., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychiatry and Psychology*, 17, 89–100. <http://dx.doi.org/10.1111/j.1469-7610.1976.tb00381.x>
- Zhai, C., & Wibowo, S. (2023). A systematic review on artificial intelligence dialogue systems for enhancing English as foreign language students' interactional competence in the university. *Computers and Education: Artificial Intelligence*, 4(2023), 100134. <https://doi.org/10.1016/j.caeai.2023.100134>
- Åkerfeldt, A., Andersson, P., Godhe, A.-L., Grönlund, Å., Heintz, F., Hård af Segerstad, Y., ... Wennås Brante, E. (2021, December 16). *Fridolin backar in i framtiden om digitala läromedel*. [Fridolin goes backwards into the future of digital teaching materials, my translation into English]. Dagens Nyheter. <http://urn.kb.se/resolve?urn=urn:nbn:se:hj:diva-55358>

### **Presentations**

- Clark, D. (2023, July 3-5). *AI changes everything!* [Keynote speaker]. EDULEARN23, Palma de Mallorca, Spain.
- Healey, D. (2022, May 31-June 4). *Critical CALL: Issues and concerns*. [Presentation on Zoom]. CALICO 2022, Seattle, USA.
- Holmes, W. (2022, June 4–6). *Artificial intelligence and education: A critical studies perspective*. [Keynote speaker]. EDULEARN22, Palma de Mallorca, Spain.
- Luckin, R. (2018, November 19). *AI in education: Blending human intelligence and machine learning*. [Keynote speaker]. GRADE Symposium, Sundsvall, Sweden.
- Teigland, R. (2018, November 13). *Walking the line between theory and practice*. [Keynote speaker]. Chalmers, Gothenburg, Sweden. <https://www.slideshare.net/eteigland/presentations>

## Videos

- Edling, S. (2019, June). *Etiska perspektiv och deras konsekvenser för individens etiska ansvar* [Virtual presentation].
- Engineered Arts. (2023, April 2). *Ameca speaks multiple languages and translates*. [Video]. YouTube clip. <https://youtu.be/wJyYMGswc1g>
- Google Duplex. (2018). *Google Duplex: A.I. assistant calls local businesses to make appointments*. [Video]. YouTube clip. <https://youtu.be/D5VN56jQMWM>
- Registre, J. (2023, February 23). *ChatGPT Voice Just KILLED Traditional Language Learning*. [Video]. YouTube clip. <https://youtu.be/J6e0JHl3C4>
- Swedish Public Service Television. (2023, March 18). *Här övar Varbergs gymnasieelever engelska med hjälp av AI* [Upper secondary school students in Varberg are practicing English with the help of AI, my translation from Swedish]. [Video]. SVT. <https://www.svt.se/nyheter/lokalt/halland/har-ovar-varbergs-gymnasieelever-engelska-med-hjalp-av-ai>

PART 2

---

**THE PAPERS**

